B116

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number
WO 01/53454 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US00/34983

(22) International Filing Date: 22 December 2000 (22.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/488,725	21 January 2000 (21.01.2000)	US
09/552,317	25 April 2000 (25.04.2000)	US
09/598,042	20 June 2000 (20.06.2000)	US
09/620,312	19 July 2000 (19.07.2000)	US
09/653,450	31 August 2000 (31.08.2000)	US
09/729,739	4 December 2000 (04.12.2000)	US
Not furnished	21 December 2000 (21.12.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	09/488,725 (CIP)
Filed on	21 January 2000 (21.01.2000)
US	09/552,317 (CIP)
Filed on	25 April 2000 (25.04.2000)
US	09/598,042 (CIP)
Filed on	20 June 2000 (20.06.2000)
US	09/620,312 (CIP)
Filed on	19 July 2000 (19.07.2000)
US	09/653,450 (CIP)
Filed on	31 August 2000 (31.08.2000)
US	09/729,739 (CIP)
Filed on	4 December 2000 (04.12.2000)
US	Not furnished (CIP)
Filed on	21 December 2000 (21.12.2000)

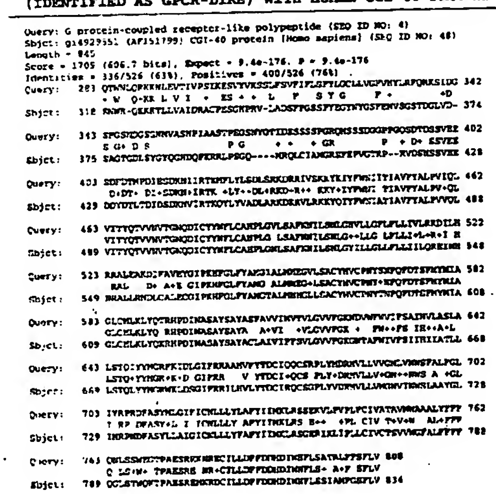
(71) Applicant *(for all designated States except US)*: **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **YAMAZAKI, Victoria** [JP/US]; 883 Portwalk Place, Redwood Shores, CA 94065 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117

[Continued on next page]

(54) Title: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES

BLASTP ALIGNMENT OF SEQ ID NO: 4, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH HUMAN CGI-40 PROTEIN SEQ ID NO: 48



(57) Abstract: The invention provides novel polynucleotides and polypeptides encoded by such polynucleotides and mutants or variants thereof that correspond to a novel human secreted GPCR-like polypeptide. These polynucleotides comprise nucleic acid sequences isolated from cDNA libraries from human leukocyte mRNA (GIBCO Laboratories) (SEQ ID NO: 1), from human adult liver mRNA (GIBCO) (SEQ ID NO: 10); from human adult kidney mRNA (GIBCO) (SEQ ID NO: 17); from human adult brain mRNA (GIBCO) (SEQ ID NO: 26) and from human adult kidney mRNA (Invitrogen) (SEQ ID NO: 33). Other aspects of the invention include vectors containing processes for producing novel human secreted GPCR-like polypeptides, and antibodies specific for such polypeptides.

WO 01/53454 A2

(US). **ZHOU, Ping** [CN/US]; 1461 Japaul Lane, San Jose, CA 95132 (US). **WANG, Dunrui** [CN/US]; 932 La Palma, Milpitas, CA 95035 (US). **ZHANG, Jie** [CN/US]; 4930 Poplar Terrace, Campbell, CA 95008 (US). **REN, Feiyan** [US/US]; 7703 Oak Meadow Court, Cupertino, CA 95014 (US). **ASUNDI, Vinod** [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). **DRMANAC, Radoje, T.** [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovsky & Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDES AND POLYNUCLEOTIDES

1. TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods. In particular, the invention relates to novel G protein-coupled receptor-like (GPCR-like) polypeptides.

2. BACKGROUND ART

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences. Proteins are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity. It is to these polypeptides and the polynucleotides encoding them that the present invention is directed. In particular, this invention is directed to novel GPCR-like polypeptides and polynucleotides.

Effective intercellular communication is obligatory for the successful survival of multicellular organisms. Environmental cues are normally recognized by a plethora of specific receptors present mainly on the cell membrane. Binding of the appropriate ligand activates the receptor, which initiates different signaling cascades that finally result in modification of cellular activity. Cells communicate with other cells, extracellular matrix, soluble hormones and chemokines, pheromones, toxins, viruses and bacteria using these receptors. The nature of the interactions and resulting signal transduction events define the fate of cell.

G protein-coupled receptors (GPCRs) constitute an evolutionarily conserved, but functionally very diverse family of such membrane receptors. All GPCR members share a common central seven transmembrane helices, termed TM-I through TM-VII; connected by three intracellular and three extracellular loops. Two conserved cysteine residues in these helices form a disulfide link that may be important for packing and stabilization of these seven TMs. The unique extracellular regions of individual GPCRs recognize specific ligands, the disulfide

bridge is implicated in interactions with agonists and antagonists, and the third intracellular loop interacts with G proteins that in turn activate second messengers such as cyclic adenosine monophosphate (cAMP), phospholipase C, inositol triphosphate, or ion channel proteins.

In vertebrates, the GPCR family contains more than 2000 gene members that can be subdivided into at least five subfamilies based on their ligand-binding properties. Family 1a binds small ligands including rhodopsin, odorants, and beta-adrenergic receptors and interestingly the ligand-binding site is contained within the seven TM region. Family 1b binds small peptides and the binding site is located in the extracellular loop and the seven TM region, while family 1c binds large glycoproteins and the binding site is mainly located in the extracellular domain with extracellular loops making some contacts. Family 2 is similar to family 1c with respect to ligand-binding but does not share any sequence similarities with family 1. Family 3 contains the Ca^{2+} sensing receptors while family 4 has pheromone receptors as its members. Finally, family 5 primarily consists of receptors involved in embryonic development. Thus, GPCRs are involved in the recognition and transduction of messages as diverse as light, Ca^{2+}, odorants, small molecules including amino acids, nucleotides, lipids, and peptides, hormones and pheromones, chemokines and complement, neurotransmitters, as well as larger proteins. GPCRs control the activity of enzymes, ion channels, and transport of vesicles via the catalysis of the GDP-GTP exchange on heterotrimeric G proteins ($G\alpha-\beta\gamma$) (Bockaert and Pin, (1999) EMBO J. 18, 1723-1729).

Olfactory GPCRs are responsible for transmission of volatile chemical signals from the environment through the olfactory neurons to the brain. Homologous receptors are also expressed in human testis and aid in sperm chemotaxis. Chemotactic GPCRs are also involved in immune response. Chemokines, platelet activating factor, and complement components all use GPCRs to transduce signals in the immune system. Regulation of GPCR activity is achieved at several levels. Apart from transcriptional and translational regulation, the GPCR family members have been shown to homo- and heterodimerize which can modulate their functions. Further, it has been shown that GPCRs can also interact with arrestins and certain PDZ domain containing proteins to transduce signals.

Abnormal GPCR function has been reported for various diseases including hyperthyroidism, familial precocious puberty, and congenital nephrogenic diabetes insipidus. Some of the GPCRs have been shown to function as proto-oncogenes and can be activated by mutagenesis. GPCRs are thus involved in many of the pathologies of human diseases.

Thus, the GPCR-like polypeptides and polynucleotides of the invention may be used in the treatment of diseases of ophthalmic, neurological, immunological, and nephritic systems.

They may also be used to treat hormonal dysfunction, cancer and other neoplasia, atherosclerosis, and diabetes.

3. SUMMARY OF THE INVENTION

This invention is based on the discovery of novel GPCR-like polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies. Specifically, the polynucleotides of the present invention are based on GPCR- like polynucleotides isolated from cDNA libraries prepared from human leukocyte mRNA (GIBCO Laboratories) (SEQ ID NO: 1), from human adult liver mRNA (GIBCO) (SEQ ID NO: 10); from human adult kidney mRNA (GIBCO) (SEQ ID NO: 17); from human adult brain mRNA (GIBCO) (SEQ ID NO: 26) and from human adult kidney mRNA (Invitrogen) (SEQ ID NO: 33).

The compositions of the present invention additionally include vectors such as expression vectors containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The compositions of the invention provide isolated polynucleotides that include, but are not limited to, a polynucleotide comprising the nucleotide sequence set forth in SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62; or a fragment of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62; a polynucleotide comprising the full length protein coding sequence of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 (for example, SEQ ID NO: 4, 13, 20, 29, 36, and 42); and a polynucleotide comprising the nucleotide sequence of the mature protein coding sequence of any of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any of the nucleotide sequences set forth in SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62; (b) a nucleotide sequence encoding any of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63; a polynucleotide which is an allelic variant of any polynucleotides recited above having at least 70% polynucleotide sequence identity to the polynucleotides; a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the

peptides recited above; or a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptide comprising SEQ ID NO: 4, 13, 20, 29, 36, or 42. Preferably the polynucleotides include a polynucleotide comprising the sequence set forth in nucleotides 1-1351 of SEQ ID NO: 19 or nucleotides 271-1351 of SEQ ID NO: 19. Preferably the polynucleotides include a polynucleotide comprising the sequence set forth in nucleotides 1-1668 of SEQ ID NO: 28, nucleotides 52-1668 of SEQ ID NO: 28, or nucleotides 2845-3993 of SEQ ID NO: 28.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or unique identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information are provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention further provides cloning or expression vectors comprising at least a fragment of the polynucleotides set forth above and host cells or organisms transformed with these expression vectors. Useful vectors include plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The compositions of the present invention include polypeptides comprising, but not limited to, an isolated polypeptide selected from the group comprising the amino acid sequence of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the protein sequences listed as SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63 and substantial equivalents thereof that retain biological or immunological activity are also contemplated. Preferably the polypeptides include a polypeptide

comprising the sequence set forth in amino acid residues 1-360 of SEQ ID NO: 20. Preferably the polypeptides include a polypeptide comprising the sequence set forth in amino acid residues 1-539 of SEQ ID NO: 29 or the sequence set forth in amino acid residues 932-1314 of SEQ ID NO: 29. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Pharmaceutical compositions of the invention may comprise a polypeptide of the invention and an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also relates to methods for producing a polypeptide of the invention comprising culturing host cells comprising an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the protein or peptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such a process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use in an array, use in computer-readable media, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of antisense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example,.a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a peptide of the present invention and a pharmaceutically acceptable carrier.

Thus, the GPCR-like polypeptides and polynucleotides of the invention may be used in the treatment of diseases of ophthalmic, neurological, immunological, and nephritic systems. They may also be used to treat hormonal dysfunction, cancer and other neoplasia, atherosclerosis, and diabetes.

The methods of the invention also provide methods for the treatment of disorders as recited herein which comprise the administration of a therapeutically effective amount of a composition comprising a polynucleotide or polypeptide of the invention and a pharmaceutically acceptable carrier to a mammalian subject exhibiting symptoms or tendencies related to disorders as recited herein. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising the step of administering a composition comprising compounds and other substances that modulate the overall activity of the target gene products and a pharmaceutically acceptable carrier. Compounds and other substances can effect such modulation either on the level of target gene/protein expression or target protein activity. Specifically, methods are provided for preventing, treating or ameliorating a medical condition, including viral diseases, which comprises administering to a mammalian subject, including but not limited to humans, a therapeutically effective amount of a composition comprising a polypeptide of the invention or a therapeutically effective amount of a composition comprising a binding partner of (e.g., antibody specifically reactive for) GPCR-like polypeptides of the invention. The mechanics of the particular condition or pathology will dictate whether the polypeptides of the invention or binding partners (or inhibitors) of these would be beneficial to the individual in need of treatment.

According to this method, polypeptides of the invention can be administered to produce an *in vitro* or *in vivo* inhibition of cellular function. A polypeptide of the invention can be administered *in vivo* alone or as an adjunct to other therapies. Conversely, protein or other active ingredients of the present invention may be included in formulations of a particular agent to minimize side effects of such an agent.

The invention further provides methods for manufacturing medicaments useful in the above-described methods.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample (e.g., tissue or sample). Such

methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions.

The invention provides a method for detecting a polypeptide of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting formation of the complex, so that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention.

The invention provides a method for identifying a compound that binds to the polypeptide of the present invention comprising contacting the compound with the polypeptide under conditions and for a time sufficient to form a polypeptide/compound complex and detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide is identified.

Also provided is a method for identifying a compound that binds to the polypeptide comprising contacting the compound with the polypeptide in a cell for a time sufficient to form a polypeptide/compound complex wherein the complex drives expression of a reporter gene sequence in the cell and detecting the complex by detecting reporter gene sequence expression so that if the polypeptide/compound complex is detected a compound that binds to the polypeptide is identified.

4. BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 3 (i.e. SEQ ID NO: 4) GPCR-like polypeptide and human CGI-40 protein (Lai et al, (2000) Genome Res. 10, 703-713) (SEQ ID NO: 48), indicating that the two

sequences share 76% similarity over 526 amino acid residues and 63% identity over the same 526 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 2 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 3 (i.e. SEQ ID NO: 4) GPCR-like polypeptide and protein of clone CT748_2 (Patent Application No. WO9824905) (SEQ ID NO: 49), indicating that the two sequences share 97% similarity over 445 amino acid residues and 96% identity over the same 445 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 3 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 12 (i.e. SEQ ID NO: 13) GPCR-like polypeptide and human six transmembrane epithelial antigen of prostate (Hubert et al, (1999) Proc. Natl. Acad. Sci. U.S.A. 96, 14523-14528) (SEQ ID NO: 50), indicating that the two sequences share 68% similarity over 267 amino acid residues and 47% identity over the same 267 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 4 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 12 (i.e. SEQ ID NO: 13) GPCR-like polypeptide and human STRAP-1 protein (Patent Application No. WO9962941) (SEQ ID NO: 51), indicating that the two sequences share 68% similarity over 267 amino acid residues and 47% identity over the same 267 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 5A, 5B, and 5C show the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 19 (i.e. SEQ ID NO: 20) GPCR-like polypeptide and rat seven transmembrane receptor protein (Abe et al, (1999) J. Biol. Chem. 274, 19957-19964) (SEQ ID NO: 52), indicating that the two sequences share 81% similarity over 1354 amino

acid residues and 72% identity over the same 1354 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 6A, and 6B show the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 19 (i.e. SEQ ID NO: 20) GPCR-like polypeptide and human brain-derived G protein-coupled receptor protein (Patent Application No. WO200008053) (SEQ ID NO: 53), indicating that the two sequences share 100% similarity over 986 amino acid residues and 100% identity over the same 986 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 7 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 28 (i.e. SEQ ID NO: 29) GPCR-like polypeptide and putative seven pass transmembrane protein (Spangenberg et al (1998) Genomics 48, 178-185) (SEQ ID NO: 54), indicating that the two sequences share 72% similarity over 323 amino acid residues and 57% identity over the same 323 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 8 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 28 (i.e. SEQ ID NO: 29) GPCR-like polypeptide and human h-TRAAK polypeptide #1 (Patent Application No. WO200026253) (SEQ ID NO: 55), indicating that the two sequences share 100% similarity over 392 amino acid residues and 100% identity over the same 392 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 9 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 35 (i.e. SEQ ID NO: 36) GPCR-like polypeptide and human olfactory

receptor (Rouquier et al, (1998) Nature Genet. 18 (3), 243-250) (SEQ ID NO: 56), indicating that the two sequences share 92% similarity over 166 amino acid residues and 87% identity over the same 166 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 10 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 35 (i.e. SEQ ID NO: 36) GPCR-like polypeptide and human G protein-coupled receptor GPR1 protein (Patent Application No. WO9630406) (SEQ ID NO: 57), indicating that the two sequences share 93% similarity over 171 amino acid residues of and 92% identity over the same 171 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 11 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 41 (i.e. SEQ ID NO: 42) GPCR-like polypeptide and human gene AC005587, similar to mouse olfactory receptor 13 polypeptide (SEQ ID NO: 58), indicating that the two sequences share 81% similarity over 304 amino acid residues and 68% identity over the same 304 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

Figure 12 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 41 (i.e. SEQ ID NO: 42) GPCR-like polypeptide and human G protein-coupled receptor GPR1 polypeptide (Patent Application No. WO9630406) (SEQ ID NO: 59), indicating that the two sequences share 91% similarity over 287 amino acid residues and 90% identity over the same 287 amino acid residues, wherein A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine. Gaps are presented as dashes.

5. DETAILED DESCRIPTION OF THE INVENTION

The GPCR-like polypeptide of SEQ ID NO: 4 is an approximately 827-amino acid transmembrane protein with a predicted molecular mass of approximately 93 kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 4 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 0.011. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 6. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 4 is homologous to human CGI-40 protein and with protein of clone CT748_2.

Figure 1 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 3 (i.e. SEQ ID NO: 4) GPCR-like polypeptide and human CGI-40 protein (Lai et al, (2000) Genome Res. 10, 703-713) (SEQ ID NO: 48), indicating that the two sequences share 76% similarity over 526 amino acid residues and 63% identity over the same 526 amino acid residues.

Figure 2 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 3 (i.e. SEQ ID NO: 4) GPCR-like polypeptide and protein of clone CT748_2 (Patent Application No. WO9824905) (SEQ ID NO: 49), indicating that the two sequences share 97% similarity over 445 amino acid residues and 96% identity over the same 445 amino acid residues.

A predicted approximately nineteen-residue signal peptide is encoded from approximately residue 1 through residue 19 of SEQ ID NO: 4 (SEQ ID NO: 7). The extracellular portion is useful on its own. This can be confirmed by expression in mammalian cells and sequencing of the cleaved product. The signal peptide region was predicted using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program.

The GPCR-like polypeptide of SEQ ID NO: 13 is an approximately 488-amino acid transmembrane protein with a predicted molecular mass of approximately 55-kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch

- search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 13 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 0.017. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 15. Protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 4 is homologous to six transmembrane epithelial antigen of prostate and with human STRAP-1 protein.

Figure 3 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 12 (i.e. SEQ ID NO: 13) GPCR-like polypeptide and human six transmembrane epithelial antigen of prostate (Hubert et al, (1999) Proc. Natl. Acad. Sci. U.S.A. 96, 14523-14528) (SEQ ID NO: 50), indicating that the two sequences share 68% similarity over 267 amino acid residues and 47% identity over the same 267 amino acid residues.

Figure 4 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 12 (i.e. SEQ ID NO: 13) GPCR-like polypeptide and human STRAP-1 protein (Patent Application No. WO9962941) (SEQ ID NO: 51), indicating that the two sequences share 68% similarity over 267 amino acid residues and 47% identity over the same 267 amino acid residues.

The GPCR-like polypeptide of SEQ ID NO: 20 is an approximately 1346-amino acid transmembrane protein with a predicted molecular mass of approximately 151-kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 20 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 2.7e-24. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 22. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 20 is homologous to the rat seven transmembrane receptor and to the human brain-derived G protein-coupled receptor proteins.

Figure 5A, 5B, and 5C show the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 19 (i.e. SEQ ID NO: 20) GPCR-like polypeptide and rat seven transmembrane receptor protein (Abe et al, (1999) J. Biol. Chem. 274, 19957-19964)

(SEQ ID NO: 52), indicating that the two sequences share 81% similarity over 1354 amino acid residues and 72% identity over the same 1354 amino acid residues.

Figure 6A, and 6B show the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 19 (i.e. SEQ ID NO: 20) GPCR-like polypeptide and human brain-derived G protein-coupled receptor protein (Patent Application No. WO200008053) (SEQ ID NO: 53), indicating that the two sequences share 100% similarity over 986 amino acid residues and 100% identity over the same 986 amino acid residues.

A predicted approximately twenty one-residue signal peptide is encoded from approximately residue 1 through residue 21 of SEQ ID NO: 20 (SEQ ID NO: 23). The extracellular portion is useful on its own. This can be confirmed by expression in mammalian cells and sequencing of the cleaved product. The signal peptide region was predicted using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program.

The GPCR-like polypeptide of SEQ ID NO: 29 is an approximately 1314-amino acid transmembrane protein with a predicted molecular mass of approximately 147-kDa unglycosylated. Hyseq's sequence database searches with the Pfam models that were categorized under G protein-coupled receptors using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 29 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 0.0036. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 31. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 29 is homologous to the putative seven pass transmembrane protein and to the human h-TRAAK polypeptide #1.

Figure 7 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 28 (i.e. SEQ ID NO: 29) GPCR-like polypeptide and putative seven pass transmembrane protein (Spangenberg et al (1998) Genomics 48, 178-185) (SEQ ID NO: 54), indicating that the two sequences share 72% similarity over 323 amino acid residues and 57% identity over the same 323 amino acid residues.

Figure 8 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 28 (i.e. SEQ ID NO: 29) GPCR-like polypeptide and human h-TRAAK polypeptide #1 (Patent Application No. WO200026253) (SEQ ID NO: 55), indicating that the

two sequences share 100% similarity over 392 amino acid residues and 100% identity over the same 392 amino acid residues.

The GPCR-like polypeptide of SEQ ID NO: 36 is an approximately 194-amino acid transmembrane protein with a predicted molecular mass of approximately 22-kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 36 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 1.8e-28. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 38. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 36 is homologous to the human olfactory receptor protein and to the human G protein-coupled receptor GPR1 polypeptide.

Figure 9 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 35 (i.e. SEQ ID NO: 36) GPCR-like polypeptide and human olfactory receptor (Rouquier et al, (1998) Nature Genet. 18 (3), 243-250) (SEQ ID NO: 56), indicating that the two sequences share 92% similarity over 166 amino acid residues and 87% identity over the same 166 amino acid residues.

Figure 10 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 35 (i.e. SEQ ID NO: 36) GPCR-like polypeptide and human G protein-coupled receptor GPR1 protein (Patent Application No. WO9630406) (SEQ ID NO: 57), indicating that the two sequences share 93% similarity over 171 amino acid residues of and 92% identity over the same 171 amino acid residues.

A predicted approximately thirty five-residue signal peptide is encoded from approximately residue 1 through residue 35 of SEQ ID NO: 36 (SEQ ID NO: 39). The extracellular portion is useful on its own. This can be confirmed by expression in mammalian cells and sequencing of the cleaved product. The signal peptide region was predicted using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program.

The GPCR-like polypeptide of SEQ ID NO: 42 is an approximately 308-amino acid transmembrane protein with a predicted molecular mass of approximately 34-kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were

categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 42 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 1.1e-47. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 44. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 42 is homologous to the mouse olfactory receptor 13 polypeptide and to the human G protein-coupled receptor GPR1 polypeptide.

Figure 11 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 41 (i.e. SEQ ID NO: 42) GPCR-like polypeptide and human gene AC005587, similar to mouse olfactory receptor 13 polypeptide (SEQ ID NO: 58), indicating that the two sequences share 81% similarity over 304 amino acid residues and 68% identity over the same 304 amino acid residues.

Figure 12 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 41 (i.e. SEQ ID NO: 42) GPCR-like polypeptide and human G protein-coupled receptor GPR1 polypeptide (Patent Application No. WO9630406) (SEQ ID NO: 59), indicating that the two sequences share 91% similarity over 287 amino acid residues and 90% identity over the same 287 amino acid residues.

A predicted approximately forty two-residue signal peptide is encoded from approximately residue 1 through residue 42 of SEQ ID NO: 42 (SEQ ID NO: 45). The extracellular portion is useful on its own. This can be confirmed by expression in mammalian cells and sequencing of the cleaved product. The signal peptide region was predicted using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program.

Thus, the GPCR-like polypeptides and polynucleotides of the invention may be used in the treatment of diseases of ophthalmic, neurological, immunological, and nephritic systems. They may also be used to treat hormonal dysfunction, cancer and other neoplasia, atherosclerosis, and diabetes.

5.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide that retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms “biologically active” or “biological activity” refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise “biologically active” or “biological activity” refers to the capability of the natural, recombinant or synthetic GPCR-like peptide, or any peptide thereof, to induce a specific biological response in appropriate animals or cells and to bind with specific antibodies. The term “GPCR-like biological activity” refers to biological activity that is similar to the biological activity of a GPCR-like polypeptide.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms “complementary” or “complementarity” refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5’-AGT-3’ binds to the complementary sequence 3’-TCA-5’. Complementarity between two single-stranded molecules may be “partial” such that only some of the nucleic acids bind or it may be “complete” such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term “embryonic stem cells (ES)” refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term “germ line stem cells (GSCs)” refers to stem cells derived from primordial stem cells that provide a steady and continuous source of germ cells for the production of gametes. The term “primordial germ cells (PGCs)” refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides that modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs

include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs is nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or “nucleic acid” or “polynucleotide” or “oligonculeotide” are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences, A is adenine, G is guanine, C is cytosine, T is thymine, and N is A, G, C, or T (U). It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequence herein may be replaced with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or “probe” or “primer” are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30 nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to a portion of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They

may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their entirety.

The nucleic acid sequences of the present invention also include the sequence information from any of the nucleic acid sequences of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62. The sequence information can be a segment of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 that uniquely identifies or represents the sequence information of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3 x 25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms "operably linked" or "operably associated" refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence.

While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term "pluripotent" refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms "polypeptide" or "peptide" or "amino acid sequence" refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 200 amino acids, more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term "translated protein coding portion" means a sequence which encodes for the full length protein which may include any leader sequence or a processing sequence.

The term "mature protein coding sequence" refers to a sequence which encodes a peptide or protein without any leader/signal sequence. The "mature protein portion" refers to that portion of the protein without the leader/signal sequence. The peptide may have the leader sequences removed during processing in the cell or the protein may have been produced synthetically or using a polynucleotide only encoding for the mature protein coding sequence. It is contemplated that the mature protein portion may or may not include the initial methionine residue. The initial methionine is often removed during processing of the peptide.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover

rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other components normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant

protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2):134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligonucleotides), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 90% sequence identity, most preferably at least 95% identity. Substantially equivalent nucleotide sequence of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, and most preferably at least about 95% identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J. (1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term “totipotent” refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

5.2 NUCLEIC ACIDS OF THE INVENTION

The invention is based on the discovery of a novel secreted GPCR-like polypeptide, the polynucleotides encoding the GPCR-like polypeptide and the use of these compositions for the diagnosis, treatment or prevention of cancers and other immunological disorders.

The isolated polynucleotides of the invention include, but are not limited to a polynucleotide comprising any of the nucleotide sequences of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62; a fragment of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62; a polynucleotide comprising the full length protein coding sequence of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 (for example SEQ ID NO: 4, 13, 20, 29, 36, or 42); and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polynucleotides of any one of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62; (b) a polynucleotide encoding any one of the

polypeptides of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63. Preferably the polynucleotides include a polynucleotide comprising the sequence set forth in nucleotides 1-1351 of SEQ ID NO: 19 or nucleotides 271-1351 of SEQ ID NO: 19. Preferably the polynucleotides include a polynucleotide comprising the sequence set forth in nucleotides 1-1668 of SEQ ID NO: 28, nucleotides 52-1668 of SEQ ID NO: 28, or nucleotides 2845-3993 of SEQ ID NO: 28. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include all of the coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, more typically at least about 90%, and even more typically at least about 95%, sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62, or complements thereof, which fragment is greater than about 5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to any one of the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided in SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor result for the nucleic acids of the present invention, including SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for

Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990))

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent

nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a

functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat; pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of

expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3

(Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intramuscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

5.3 ANTISENSE

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO:SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term

"noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO:SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 , antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of a mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of a mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of a mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

5.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as a mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave a mRNA transcripts to thereby inhibit translation of a mRNA. A ribozyme having specificity for a nucleic acid of the invention can be designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO:SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62). For

example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a SECX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may

combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl AcidRes* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

5.5 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of GPCR-like DNA sequences allows for modification of cells to permit, or increase, expression of GPCR-like polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased GPCR-like polypeptide expression by replacing, in whole or in part, the naturally occurring GPCR-like promoter with all or part of a heterologous promoter so that the cells GPCR-like polypeptide is expressed at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to GPCR-like encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the GPCR-like coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the GPCR-like coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE, dextran-mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site. splice donor and acceptor sites, transcriptional termination sequences, and 5′ flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli, Bacillus subtilis, Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from

a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

5.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequence set forth as any one of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of the nucleotide sequences set forth in SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63 or the corresponding full length or mature protein; and "substantial equivalents" thereof (e.g., with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, typically at least about 95%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63. Preferably the polypeptides include a polypeptide comprising the sequence set forth in amino acid residues 1-360 of SEQ ID NO: 20. Preferably the polypeptides include a polypeptide comprising the sequence set forth in amino acid residues 1-539 of SEQ ID NO: 29 or the sequence set forth in amino acid residues 932-1314 of SEQ ID NO: 29.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding

sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which it is expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: *A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63.

The protein of the invention may also be expressed as a product of transgenic animals, *e.g.*, as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, *e.g.*, U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein

may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP-HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). The polypeptides of the invention include GPCR-like analogs. This embraces fragments of GPCR-like polypeptide of the invention, as well GPCR-like polypeptides which comprise one or more amino acids deleted, inserted, or substituted. Also, analogs of the GPCR-like polypeptide of the invention embrace fusions of the GPCR-like polypeptides or modifications of the GPCR-like polypeptides, wherein the GPCR-like polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the GPCR-like polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to neurons, e.g., antibodies to central nervous system, or antibodies to receptor and ligands expressed on neuronal cells. Other moieties

which may be fused to GPCR-like polypeptide include therapeutic agents which are used for treatment, for example anti-depressant drugs or other medications for neurological disorders. Also, GPCR-like polypeptides may be fused to neuron growth modulators, and other chemokines for targeted delivery.

5.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), the eMatrix software (Wu et al., J. Comp. Biol., vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, vol 4, pp. 202-209, herein incorporated by reference), the GeneAtlas software (Molecular Simulations Inc. (MSI), San Diego, CA) (Sanchez and Sali (1998) Proc. Natl. Acad. Sci., 95, 13597-13602; Kitson DH et al, (2000) "Remote homology detection using structural modeling – an evaluation" Submitted; Fischer and Eisenberg (1996) Protein Sci. 5, 947-955), Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark), Pfam, which are multiple protein sequence alignment and hidden Markov models of common protein domains (Wang et al (2000) submitted and Bateman et al (2000) Nucleic Acid Res. 28, 263-266) and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

5.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to another polypeptide. Within a fusion protein the polypeptide according to the invention can

correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide(s) according to the invention and the other polypeptide(s) are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus or in the middle.

For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein. In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e,g.,* cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used to bind and to dimerize 2 receptors and thereby transduce an intracellular signal. The immunoglobulin fusion proteins may also be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using

anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

5.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or

both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

5.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model

systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies; of animals that fail to express functional GPCR-like polypeptide or that express a variant of GPCR-like polypeptide. Such animals are useful as models for studying the *in vivo* activities of GPCR-like polypeptide as well as for studying modulators of the GPCR-like polypeptide.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter

can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

5.10 USES AND BIOLOGICAL ACTIVITY OF HUMAN GPCR-LIKE POLYPEPTIDE

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

5.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify

potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

The polypeptides of the invention are also useful for making antibody substances that are specifically immunoreactive with GPCR-like proteins. Antibodies and portions thereof (e.g., Fab fragments) which bind to the polypeptides of the invention can be used to identify the presence of such polypeptides in a sample. Such determinations are carried out using any suitable immunoassay format, and any polypeptide of the invention that is specifically bound by the antibody can be employed as a positive control.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

5.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Additionally, the polypeptides of the invention can be used as molecular weight markers, and as a food supplement. A polypeptide consisting of SEQ ID NO: 4, for example, has a molecular mass of approximately 93 kDa in its unprocessed and unglycosylated state. Protein food supplements are well known and the formulation of suitable food supplements including polypeptides of the invention is within the level of skill in the food preparation art.

5.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7,

Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin-γ, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9-Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

5.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* may maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder layer for the stem cell populations in culture or in vivo. Stromal support cells

for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

5.10.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines, including those assays cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

5.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic

disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pp. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

5.10.7 IMMUNE FUNCTION STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders

(including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses

or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a therapeutic composition of the invention may prevent cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, a lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular therapeutic compositions in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of therapeutic compositions of the invention on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of

autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block stimulation of T cells can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (e.g., a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response may be useful in cases of viral infection, including systemic viral diseases such as influenza, the common cold, and encephalitis.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC

class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

5.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example,

U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

5.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan,

A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

5.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

5.10.11 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology

including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide, Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX),

Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

5.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

5.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and

the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science 282*:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the

binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

5.10.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecule, that modulate (*i.e.,* increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

5.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions, including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflamation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic mylegenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

5.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

5.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of

therapeutic utility, include but are not limited to nervous system injuries, and diseases or disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

5.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part

size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

5.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled

nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

5.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

5.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have

numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

5.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the GPCR-like polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of GPCR-like polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1μg/kg to 10 mg/kg of patient body weight. For parenteral administration, GPCR-like polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

5.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not

interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing,

prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co- administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

5.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody,

targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

5.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other

glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be

provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active

ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for use in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight)

copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredient of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredient of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic

assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

5.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no

toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

5.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

5.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 4, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of -related protein that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for

example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

5.13.1 Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known

techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

5.13.2 Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas

typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors,

which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

5.13.2 Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

5.13.3 Human Antibodies

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al,(Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a

mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

5.13.4 F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., *1989* Science *246*: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or

derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

5.13.5 Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid

side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The

fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), Fc γRII (CD32) and Fc γRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

5.13.6 Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

5.13.7 Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

5.13.8 Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238:

1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

5.14 COMPUTER READABLE SEQUENCES

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer

readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to,

Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids, or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

5.15 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

5.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Techniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay

format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

5.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

5.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present

invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

5.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from of any of the nucleotide sequences SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of chromosome spreads has been described, among other places, in Verma et

al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

5.20 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, 1990 J. Clin Microbiol 28(6) 1462-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, Mol. Cell Probes 1989 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci USA 91(8) 3072-6 describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound

to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal Biochem 198(1) 138-42.

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., 1991). In this technology, a phosphoramidate bond is employed (Chu et al., 1983 Nucleic Acids 11(18) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-MeIm7), is then added to a final concentration of 10 mM 1-MeIm7. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide(EDC), dissolved in 10 mM 1-MeIm7, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res.

19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal Biochem 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) Proc. Natl. Acad. Sci USA 91(11) 5022-6. These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

5.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6. In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *CviJI*, described by Fitzgerald *et al.* (1992) Nucleic Acids

Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *CviJI* normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*CviJI***), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *CviJI*** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *CviJI*** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed).

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

5.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm^2, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 patients. For each patient, the

amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane. Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

6.0 EXAMPLES

EXAMPLE 1

Isolation of SEQ ID NO: 1, 10, 17, 26, and 33 from cDNA Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from human leukocyte mRNA (GIBCO Laboratories) (SEQ ID NO: 1), from human adult liver mRNA (GIBCO) (SEQ ID NO: 10); from human adult kidney mRNA (GIBCO) (SEQ ID NO: 17); from human adult brain mRNA (GIBCO) (SEQ ID NO: 26) and from human adult kidney mRNA (Invitrogen) (SEQ ID NO: 33) using standard PCR, sequencing by hybridization sequence signature analysis, and Sanger sequencing techniques. The inserts of the libraries

were amplified with PCR using primers specific for vector sequences flanking the inserts. These samples were spotted onto nylon membranes and interrogated with oligonucleotide probes to give sequence signatures. The clones were clustered into groups of similar or identical sequences, and single representative clones were selected from each group for gel sequencing. The 5' sequence of the amplified inserts was then deduced using the reverse M13 sequencing primer in a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single-pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer. The insert was identified as a novel sequence not previously obtained from this library and not previously reported in public databases. These sequences were designated as SEQ ID NO: 1, 10, 17, 26, and 33.

EXAMPLE 2

ASSEMBLAGE OF SEQ ID NO: 2, 11, 18, 27, 34, 60, and 62

The nucleic acids of the present invention, designated as SEQ ID NO: 2, 11, 18, 27, and 34 were assembled using SEQ ID NO: 1,10, 17, 26, and 33 as a seed, respectively. Then a recursive algorithm was used to extend the seed into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

The nearest neighbor result for the assembled contigs were obtained by a FASTA version 3 search against Genpept release 114, 117, or 118 using FASTXY algorithm. FASTXY is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for each assemblage from Genpept (and contains the translated amino acid sequences for which the assemblage encodes). The nearest neighbor results are set forth below:

SEQ ID NO:	Accession No.	Description	Smith-Waterman Score	% Identity
2	AF151799	Homo sapiens CGI-40 protein	424	71.605

11	AC005053	Homo sapiens match to ESTs AA316181 (NID: g3165221), AA032221 (NID: g1502183), and AI167942 (NID: g3701112)	881	49.580
18	AB018301	Homo sapiens KIAA0758 protein	6455	99.595
27	AF027826	Homo sapiens putative seven pass transmembrane protein	487	46.691
34	AC005587	Homo sapiens similar to mouse olfactory receptor 13, similar to P34984 (PID: g464305)	1450	71.382

Polypeptides were predicted to be encoded by SEQ ID NO: 2, 11, 18, 27, and 34 as set forth below. The polypeptides were predicted using a software program called FASTY (available from http://fasta.bioch.virginia.edu) which selects a polypeptide based on a comparison of translated novel polynucleotide to known polypeptides (W.R. Pearson, Methods in Enzymology, 183: 63-98 (1990), herein incorporated by reference).

Predicted beginning nucleotide location correspond-ing to first amino acid residue of amino acid segment	Predicted end nucleotide location correspond-ing to last amino acid residue of amino acid segment	Amino acid composition of the polypeptide encoded, wherein, (A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop Codon, /=possible nucleotide deletion, \=possible nucleotide insertion)
66	453	VGEPYIDWDEFPELLSRTAVRARKIPISDTI*KTK AKQVVKLLSNIRSQAVGILMSSLHLDMKDIQHA VVNLDNSVVDLETLQALYENRAQSDELE*IEKHG

		RSSKDKENAKSLDKPEQLYFLRFLYE (SEQ ID NO: 9)
23	462	ARGGRLRWRRLDDCLSAAESDTVAYEDLSEDYT QKKWKGLALSQRALHWNMMLENDRSMASLAG RNMMESSELTPKQEIFKGSESSNSTSGGLFGVVP GGTETGDVCEDTFKELEGQPSNEEGSRLESDFLEI IDEDKKKSTKDRY (SEQ ID NO: 16)
166	4280	ASDQSGSQPGDHSAGQANQLKLEDMKSPRRTTL CLMFIVIYSSKAALNWNYESTIHPLSLHEHEPAGE EALRQKRAVATKSPTAEEYTVNIEISFENASFLDP IKAYLNSLSFPIHGNNTDQITDILSINVTTVCRPAG NEIWCSCETGYGWPRERCLHNLICQERDVFLPGH HCSCLKELPPNGPFCLLQEDVTLNMRVRLNVGF QEDLMNTSSALYRSYKTDLETAFRKGYGILPGFK GVTVTGFKSGSVVVTYEVKTTPPSLELIHKANEQ VVQSLNQTYKMDYNSFQAVTINESNFFVTPEIIFE GDTVSLVCEKEVLSSNVSWRYEEQQLEIQNSSRF SIYTALFNNMTSVSKLTIHNITPGDAGEYVCKLIL DIFEYECKKKIDVMPIQILANEEMKVMCDNNPVS .LNCCSQGNVNWSKVEWKQEGKINIPGTPETDIDS SCSRYTLKADGTQCPSGSSGTTVIYTCEFISAYGA RGSANIKVTFISVANLTITPDPISVSEGQNFSIKCIS DVSNYDEVYWNTSAGIKIYQRFYTTRRYLDGAE SVLTVKTSTREWNGTYHCIFRYKNSYSIATKDVI VHPLPLKLNIMVDPLEATVSCSGSHHIKCCIEEDG DYKVTFHMGSSSLPAAKEVNKKQVCYKHNFNA SSVSWCSKTVDVCCHFTNAANNSVWSPSMKLNL VPGENITCQDPVIGVGEPGKVIQKLCRFSNVPSSP EE/SPLGGTITYKCVGSQWG\EKRNDCISAPINSLL QMAKALIKSPSQDEMLPTYLKDLSISIDKAEHEIS SSPGSLGAIINILDLLSTVPTQVNSEMMTHVLSTV NVILGKPVLNTWKVLQQQWTNQSSQLLHSVERF SQALQSGDSPPLSFSQTNVQMSSTVIKSSHPETYQ QRFVFPYFDLWGNVVIDKSYLENLQSDSSIVTMA FPTLQAILAQDIQENNFAESLVMTTTVSHNTTMP

		FRISMTFKNNSPSGGETKCVFWNFRLANNTGGW DSSGCYVEEGDGDNVTCICDHLTSFSILMSPDSPD PSSLLGILLDIISYVGVGFSILSLAACLVVEAVVW KSVTKNRTSYMRHTCIVNIAASLL\VANTWFIGV AAIQDNRYILCKTACVAATFFIHFFYLSVFFWML TLGLMLFYRLVFILHETSRSTQKAIAFCLGYGCPL AISVITLGATQPREVYTRKNVCWLNWEDTKALL AFAIPALIIVVVNITITIVVITKILRPSIGDKPCKQEK SSLFQISKSIGVLTPLLGLTWGFGLTTVFPGTNLV FHIIFAILNVFQGLFILLFGCLWDLKVQEALLNKF SLSRWSSQHSKSTSLGSSTPVFSMSSPISRRFNNLF GKTGTYNVSTPEATSSSLENSSSASSLLN (SEQ ID NO: 25)
1009	1208	VRGLGPRLPVFPKGKGLSVEEGGLSATTSFLLSA PSPSLHPAIPTP\RIYFPGPADSPSLSV/SRDSGLPPL TWRVTCLGLVACLPGLVPALPPAVTLGLTAAYT TLYALLFFSVYAQLWLVL\RMGHKRLS\YQT\VFL ALCL\FW\APLR\TTFFSF*FPKILPAPNN/SWGPLPF WLLYCCPVCLQFFTLT\LMNLYFA\QVVFKA/KSE ASGPKMSRGLLAVRGAFVGASLLFLLVNVLCAV L/VPCGAAAQPWALLLVRVLVSDSLFVICALSLA ACLFLCRQAGALH*HLPGGQGRAAALMPRCLLG LSAAVLRV*RTAAERPKRHLGISAAALPWPPGRC (SEQ ID NO: 32)
1206	2266	RHLLTIFHKLKIYKTINKIDFKKKRVTQLLVFCLF LCLFFSSEMVKNQTMVTEFLLLGFLLGPRIQMLL FGLFSLFYVFTLLGNGTILGLISLDSRLHTPMYFFL SHLAVVNIAYACNTVPQMLVNLLHPAKPISFAGC MT*TFLFLSFAHTECLLLVLMSYDRYVAICHPLR YFIIMTWKVCITLAITSWTCGSLLAMVHVSLILRL PFCGPREINHFFCEILSVLRLACADTWLNQVVIFA ACMFILVGPLCLVLVSYSHILAAILRIQSGEGRRK AFSTCSSHLCVVGLFFGSAIVMYMAPKSRHPEEQ QKVLFLFYSSFNPMLNPLIYNLRNVEVKGALRRA LCKESHS (SEQ ID NO: 47)

EXAMPLE 3

ASSEMBLAGE OF SEQ ID NO: 4, 13, 20, 29, 36, or 42

Using PHRAP (Univ. of Washington), full-length gene cDNA sequences and its corresponding protein sequences were generated from the assemblage of SEQ ID NO: 2, 11, 18, 27, and 34. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank. (i.e. Genpept release 117 or 119). Other computer programs, which may have been used in the editing process, were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and cg-zip-2 (Hyseq, Inc.).

A polypeptide (SEQ ID NO: 4) was predicted to be encoded by SEQ ID NO: 3 as set forth below. The polypeptide was predicted using a software program called BLASTX which selects a polypeptide based on a comparison of translated novel polynucleotide to known polynucleotides. The initial methionine starts at position 1 of SEQ ID NO: 3 and the putative stop codon, TGA, begins at position 2482 of the nucleotide sequence.

The GPCR-like polypeptide of SEQ ID NO: 4 is an approximately 827-amino acid transmembrane protein with a predicted molecular mass of approximately 93-kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 4 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 0.011. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 6. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 4 is homologous to human CGI-40 protein and with protein of clone CT748_2.

Figure 1 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 3 (i.e. SEQ ID NO: 4) GPCR-like polypeptide and human CGI-40 protein (Lai et al, (2000) Genome Res. 10, 703-713) (SEQ ID NO: 48), indicating that the two

sequences share 76% similarity over 526 amino acid residues and 63% identity over the same 526 amino acid residues.

Figure 2 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 3 (i.e. SEQ ID NO: 4) GPCR-like polypeptide and protein of clone CT748_2 (Patent Application No. WO9824905) (SEQ ID NO: 49), indicating that the two sequences share 97% similarity over 445 amino acid residues and 96% identity over the same 445 amino acid residues.

A predicted approximately nineteen-residue signal peptide is encoded from approximately residue 1 through residue 19 of SEQ ID NO: 4 (SEQ ID NO: 7). The extracellular portion is useful on its own. This can be confirmed by expression in mammalian cells and sequencing of the cleaved product. The signal peptide region was predicted using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program.

A polypeptide (SEQ ID NO: 13) was predicted to be encoded by SEQ ID NO: 12 as set forth below. The polypeptide was predicted using a software program called BLASTX which selects a polypeptide based on a comparison of translated novel polynucleotide to known polynucleotides. The initial methionine starts at position 135 of SEQ ID NO: 12 and the putative stop codon, TGA, begins at position 1599 of the nucleotide sequence.

The GPCR-like polypeptide of SEQ ID NO: 13 is an approximately 488-amino acid transmembrane protein with a predicted molecular mass of approximately 55-kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 13 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 0.017. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 15. Protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 4 is homologous to six transmembrane epithelial antigen of prostate and with human STRAP-1 protein.

Figure 3 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 12 (i.e. SEQ ID NO: 13) GPCR-like polypeptide and human six transmembrane epithelial antigen of prostate (Hubert et al, (1999) Proc. Natl. Acad. Sci.

U.S.A. 96, 14523-14528) (SEQ ID NO: 50), indicating that the two sequences share 68% similarity over 267 amino acid residues and 47% identity over the same 267 amino acid residues.

Figure 4 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 12 (i.e. SEQ ID NO: 13) GPCR-like polypeptide and human STRAP-1 protein (Patent Application No. WO9962941) (SEQ ID NO: 51), indicating that the two sequences share 68% similarity over 267 amino acid residues and 47% identity over the same 267 amino acid residues.

A polypeptide (SEQ ID NO: 20) was predicted to be encoded by SEQ ID NO: 19 as set forth below. The polypeptide was predicted using a software program called BLASTX which selects a polypeptide based on a comparison of translated novel polynucleotide to known polynucleotides. The initial methionine starts at position 272 of SEQ ID NO: 19 and the putative stop codon, TAA, begins at position 4310 of the nucleotide sequence.

The GPCR-like polypeptide of SEQ ID NO: 20 is an approximately 1346-amino acid transmembrane protein with a predicted molecular mass of approximately 151-kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 20 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 2.7e-24. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 22. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 20 is homologous to the rat seven transmembrane receptor and to the human brain-derived G protein-coupled receptor proteins.

Figure 5A, 5B, and 5C show the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 19 (i.e. SEQ ID NO: 20) GPCR-like polypeptide and rat seven transmembrane receptor protein (Abe et al, (1999) J. Biol. Chem. 274, 19957-19964) (SEQ ID NO: 52), indicating that the two sequences share 81% similarity over 1354 amino acid residues and 72% identity over the same 1354 amino acid residues.

Figure 6A, and 6B show the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 19 (i.e. SEQ ID NO: 20) GPCR-like polypeptide and human brain-derived G protein-coupled receptor protein (Patent Application No. WO200008053) (SEQ ID

NO: 53), indicating that the two sequences share 100% similarity over 986 amino acid residues and 100% identity over the same 986 amino acid residues.

A predicted approximately twenty one-residue signal peptide is encoded from approximately residue 1 through residue 21 of SEQ ID NO: 20 (SEQ ID NO: 23). The extracellular portion is useful on its own. This can be confirmed by expression in mammalian cells and sequencing of the cleaved product. The signal peptide region was predicted using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program.

SEQ ID NO: 60 is very similar to SEQ ID NO: 19. A polypeptide (SEQ ID NO: 61) was predicted to be encoded by SEQ ID NO: 60.. The initial methionine starts at 272 of SEQ ID NO: 60 and the putative stop codon begins at position 4310.

A polypeptide (SEQ ID NO: 29) was predicted to be encoded by SEQ ID NO: 28 as set forth below. The polypeptide was predicted using a software program called BLASTX which selects a polypeptide based on a comparison of translated novel polynucleotide to known polynucleotides. The initial methionine starts at position 52 of SEQ ID NO: 28 and the putative stop codon, TAA, begins at position 3994 of the nucleotide sequence.

The GPCR-like polypeptide of SEQ ID NO: 29 is an approximately 1314-amino acid transmembrane protein with a predicted molecular mass of approximately 147-kDa unglycosylated. Hyseq's sequence database searches with the Pfam models that were categorized under G protein-coupled receptors using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 29 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 0.0036. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 31. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 29 is homologous to the putative seven pass transmembrane protein and to the human h-TRAAK polypeptide #1.

Figure 7 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 28 (i.e. SEQ ID NO: 29) GPCR-like polypeptide and putative seven pass transmembrane protein (Spangenberg et al (1998) Genomics 48, 178-185) (SEQ ID NO: 54), indicating that the two sequences share 72% similarity over 323 amino acid residues and 57% identity over the same 323 amino acid residues.

Figure 8 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 28 (i.e. SEQ ID NO: 29) GPCR-like polypeptide and human h-TRAAK polypeptide #1 (Patent Application No. WO200026253) (SEQ ID NO: 55), indicating that the two sequences share 100% similarity over 392 amino acid residues and 100% identity over the same 392 amino acid residues.

A polypeptide (SEQ ID NO: 36) was predicted to be encoded by SEQ ID NO: 35 as set forth below. The polypeptide was predicted using a software program called BLASTX which selects a polypeptide based on a comparison of translated novel polynucleotide to known polynucleotides. The initial methionine starts at position 833 of SEQ ID NO: 35 and the putative stop codon, TAA, begins at position 1415 of the nucleotide sequence.

The GPCR-like polypeptide of SEQ ID NO: 36 is an approximately 194-amino acid transmembrane protein with a predicted molecular mass of approximately 22-kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 36 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 1.8e-28. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 38. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 36 is homologous to the human olfactory receptor protein and to the human G protein-coupled receptor GPR1 polypeptide.

Figure 9 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 35 (i.e. SEQ ID NO: 36) GPCR-like polypeptide and human olfactory receptor (Rouquier et al, (1998) Nature Genet. 18 (3), 243-250) (SEQ ID NO: 56), indicating that the two sequences share 92% similarity over 166 amino acid residues and 87% identity over the same 166 amino acid residues.

Figure 10 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 35 (i.e. SEQ ID NO: 36) GPCR-like polypeptide and human G protein-coupled receptor GPR1 protein (Patent Application No. WO9630406) (SEQ ID NO: 57), indicating that the two sequences share 93% similarity over 171 amino acid residues of and 92% identity over the same 171 amino acid residues.

A predicted approximately thirty five-residue signal peptide is encoded from approximately residue 1 through residue 35 of SEQ ID NO: 36 (SEQ ID NO: 39). The extracellular portion

is useful on its own. This can be confirmed by expression in mammalian cells and sequencing of the cleaved product. The signal peptide region was predicted using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program.

A polypeptide (SEQ ID NO: 42) was predicted to be encoded by SEQ ID NO: 41 as set forth below. The polypeptide was predicted using a software program called BLASTX which selects a polypeptide based on a comparison of translated novel polynucleotide to known polynucleotides. The initial methionine starts at position 485 of SEQ ID NO: 41 and the putative stop codon, TAA, begins at position 1409 of the nucleotide sequence.

The GPCR-like polypeptide of SEQ ID NO: 42 is an approximately 308-amino acid transmembrane protein with a predicted molecular mass of approximately 34-kDa unglycosylated. Hyseq's sequence database searches using the Pfam models that were categorized under G protein-coupled receptors and using the hmmsearch program (hmmsearch - search a sequence database with a profile HMM: HMMER 2.1.1 (Dec 1998) Washington University School of Medicine), SEQ ID NO 42 was found to be homologous to G protein-coupled receptor model sequences with an E-value of 1.1e-47. The homologous sequence identified using Pfam hmmsearch is shown in SEQ ID NO: 44. Further analyses with protein database searches with the BLASTP algorithm (Altschul S.F. et al., J. Mol. Evol. 36:290-300 (1993) and Altschul S.F. et al., J. Mol. Biol. 21:403-10 (1990), herein incorporated by reference) indicate that SEQ ID NO: 42 is homologous to the mouse olfactory receptor 13 polypeptide and to the human G protein-coupled receptor GPR1 polypeptide.

Figure 11 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 41 (i.e. SEQ ID NO: 42) GPCR-like polypeptide and human gene AC005587, similar to mouse olfactory receptor 13 polypeptide (SEQ ID NO: 58), indicating that the two sequences share 81% similarity over 304 amino acid residues and 68% identity over the same 304 amino acid residues.

Figure 12 shows the BLASTP amino acid sequence alignment between the protein encoded by SEQ ID NO: 41 (i.e. SEQ ID NO: 42) GPCR-like polypeptide and human G protein-coupled receptor GPR1 polypeptide (Patent Application No. WO9630406) (SEQ ID NO: 59), indicating that the two sequences share 91% similarity over 287 amino acid residues and 90% identity over the same 287 amino acid residues.

A predicted approximately forty two-residue signal peptide is encoded from approximately residue 1 through residue 42 of SEQ ID NO: 42 (SEQ ID NO: 45). The extracellular portion

is useful on its own. This can be confirmed by expression in mammalian cells and sequencing of the cleaved product. The signal peptide region was predicted using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). One of skill in the art will recognize that the actual cleavage site may be different than that predicted by the computer program.

SEQ ID NO: 62 is similar to SEQ ID NO: 35 and 41. A polypeptide (SEQ ID NO: 63) was predicted to be encoded by SEQ ID NO: 62. The initial methionine starts at 1257 of SEQ ID NO: 62 and the putative stop codon begins at position 2187.

EXAMPLE 4

A. Expression of SEQ ID NO: 4, 13, 20, 29, 36, 42, 61, or 63 in cells

Chinese Hamster Ovary (CHO) cells or other suitable cell types are grown in DMEM (ATCC) and 10% fetal bovine serum (FBS) (Gibco) to 70% confluence. Prior to transfection the media is changed to DMEM and 0.5% FCS. Cells are transfected with cDNAs for SEQ ID NO: 4, 13, 20, 29, 36, 42, 61, or 63 or with pBGal vector by the FuGENE-6 transfection reagent (Boehringer). In summary, 4 μl of FuGENE-6 is diluted in 100 μl of DMEM and incubated for 5 minutes. Then, this is added to 1 μg of DNA and incubated for 15 minutes before adding it to a 35 mm dish of CHO cells. The CHO cells are incubated at 37°C with 5% CO_2. After 24 hours, media and cell lysates are collected, centrifuged and dialyzed against assay buffer (15 mM Tris pH 7.6, 134 mM NaCl, 5 mM glucose, 3 mM $CaCl_2$ and $MgCl_2$.

B. Expression Study Using SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62

The expression of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 in various tissues is analyzed using a semi-quantitative polymerase chain reaction-based technique. Human cDNA libraries are used as sources of expressed genes from tissues of interest (adult bladder, adult brain, adult heart, adult kidney, adult lymph node, adult liver, adult lung, adult ovary, adult placenta, adult rectum, adult spleen, adult testis, bone marrow, thymus, thyroid gland, fetal kidney, fetal liver, fetal liver-spleen, fetal skin, fetal brain, fetal leukocyte and macrophage). Gene-specific primers are used to amplify portions of SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 sequence from the samples. Amplified products are separated on an agarose gel, transferred and chemically linked to a nylon filter. The filter is then hybridized with a radioactively labeled (^{33}P-dCTP) double-stranded probe generated from SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21,

26-28, 30, 33-35, 37, 41, 43, 60, or 62 using a Klenow polymerase, random-prime method. The filters are washed (high stringency) and used to expose a phosphorimaging screen for several hours. Bands indicate the presence of cDNA including SEQ ID NO: 1-3, 5, 10-12, 14, 17-19, 21, 26-28, 30, 33-35, 37, 41, 43, 60, or 62 sequences in a specific library, and thus mRNA expression in the corresponding cell type or tissue.

CLAIMS

WE CLAIM:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 2-3, 5, 11-12, 14, 18-19, 21, 27-28, 30, 34-35, 37, 41, 43, 60, or 62, the translated protein coding portion thereof, the mature protein coding portion thereof, the extracellular portion thereof, or the active domain thereof.

2. An isolated polynucleotide encoding a polypeptide with biological activity, which polynucleotide hybridizes to the complement of a polynucleotide of claim 1 under stringent hybridization conditions.

3. An isolated polynucleotide encoding a polypeptide with biological activity, said polynucleotide having greater than about 90% sequence identity with the polynucleotide of claim 1.

4. The polynucleotide of claim 1 which is a DNA sequence.

5. An isolated polynucleotide which comprises the complement of the polynucleotide of claim 1.

6. A vector comprising the polynucleotide of claim 1.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell genetically engineered to express the polynucleotide of claim 1.

9. The host cell of claim 8 wherein the polynucleotide is in operative association with a regulatory sequence that controls expression of the polynucleotide in the host cell.

10. An isolated polypeptide comprising an amino acid sequence which is at least 80% identical to the amino acid sequence selected from the group consisting of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63, the translated protein coding portion thereof, the mature protein coding portion thereof, the extracellular portion thereof, or the active domain thereof.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. A polypeptide, having GPCR-like activity, comprising at least ten consecutive amino acids from the polypeptide sequences selected from the group consisting of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63.

13. The polypeptide of claim 12, comprising at least five consecutive amino acids from the polypeptide sequences selected from the group consisting of SEQ ID NO: 4, 6-9, 13, 15-16, 20, 22-25, 29, 31-32, 36, 38-40, 42, 44-47, 61, or 63.

14. A polynucleotide encoding a polypeptide according to claim 12.

15. A polynucleotide encoding a polypeptide according to claim 13.

16. A polynucleotide encoding a polypeptide according to claim 10.

17. An antibody specific for the polypeptide of claim 10.

18. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

19. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

20. The method of claim 19, wherein the polynucleotide comprises an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

21. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

22. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions and for a time sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

23. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

24. A method of producing a GPCR-like polypeptide, comprising,

a) culturing the host cell of claim 8 under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

25. A kit comprising the polypeptide of claim 10.

26. A nucleic acid array comprising the polynucleotide of claim 1 or a unique segment of the polynucleotide of claim 1 attached to a surface.

27. The array of claim 26, wherein the array detects full-matches to the polynucleotide or a unique segment of the polynucleotide of claim 1.

28. The array of claim 26, wherein the array detects mismatches to the polynucleotide or a unique segment of the polynucleotide of claim 1.

29. A method of treatment of a subject in need of enhanced activity or expression of GPCR-like polypeptide of claim 10 comprising administering to the subject a composition selected from the group consisting of:

(a) a therapeutic amount of a agonist of said polypeptide;

(b) a therapeutic amount of the polypeptide; and

(c) a therapeutic amount of a polynucleotide encoding the polypeptide in a form and under conditions such that the polypeptide is produced,

and a pharmaceutically acceptable carrier.

30. A method of treatment of a subject having need to inhibit activity or expression of GPCR-like polypeptide of claim 10 comprising administering to the subject a composition selected from the group consisting of:

(a) a therapeutic amount of an antagonist to said polypeptide;

(b) a therapeutic amount of a polynucleotide that inhibits the expression of the nucleotide sequence encoding said polypeptide; and

(c) a therapeutic amount of a polypeptide that competes with the GPCR-like polypeptide for its ligand

and a pharmaceutically acceptable carrier.

BLASTP ALIGNMENT OF SEQ ID NO: 4, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH HUMAN CGI-40 PROTEIN SEQ ID NO: 48

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 4)
Sbjct: gi4929551 (AF151799) CGI-40 protein [Homo sapiens] (SEQ ID NO: 48)
Length = 845
Score = 1709 (606.7 bits), Expect = 9.4e-176, P = 9.4e-176
Identities = 336/526 (63%), Positives = 400/526 (76%)
Query:  283 QTWNLQRKKNLEVTIVPSIKESVYVKSSLFSVFIFLSFYLGCLLVGFVHYLRFQRKSIDG 342
            + W  Q+KK L V I  +  ES + +  L   F   S Y G    F +      +D
Sbjct:  318 ENWR-QKKKTLLVAIDRACPESGHPRV-LADSFPGSSPYEGYNYGSFENVSGSTDGLVD- 374

Query:  343 SFGSNDGSGNMVASHPIAASTPEGSNYGTIDESSSSPGRQMSSSDGGPPGQSDTDSSVEE 402
            S G+ D S            P G      + +    + GR           P  + D+ SSVEE
Sbjct:  375 SAGTGDLSYGYQGHDQFKRRLPSGQ----MRQLCIAMGRSFEPVGTRP--RVDSMSSVEE 428

Query:  403 SDFDTMPDIESDKNIIRTKMFLYLSDLSRKDRRIVSKKYKIYFWNIITIAVFYALPVIQL 462
             D+DT+ DI+SDKN+IRTK +LY++DL+RKD+R++ KKY+IYFWNI TIAVFYALPV+QL
Sbjct:  429 DDYDTLTDIDSDKNVIRTKQYLYVADLARKDKRVLRKKYQIYFWNIATIAVFYALPVVQL 488

Query:  463 VITYQTVVNVTGNQDICYYNFLCAHPLGVLSAFNNILSNLGHVLLGFLFLLIVLRRDILH 522
            VITYQTVVNVTGNQDICYYNFLCAHPLG LSAFNNILSNLG++LLG LFLLI+L+R+I H
Sbjct:  489 VITYQTVVNVTGNQDICYYNFLCAHPLGNLSAFNNILSNLGYILLGLLFLLIILQREINH 548

Query:  523 RRALEAKDIFAVEYGIPKHFGLFYAMGIALMMEGVLSACYHVCPNYSNFQFDTSFMYMIA 582
             RAL  D+ A+E GIPKHFGLFYAMG ALMMEG+LSACYHVCPNY+NFQFDTSFMYMIA
Sbjct:  549 NRALLRNDLCALECGIPKHFGLFYAMGTALMMEGLLSACYHVCPNYTNFQFDTSFMYMIA 608

Query:  583 GLCMLKLYQTRHPDINASAYSAYASFAVVIMVTVLGVVFGKNDVWFWVIFSAIHVLASLA 642
            GLCMLKLYQ RHPDINASAYSAYA  A+VI  +VLGVVFGK +  FW++FS IH++A+L
Sbjct:  609 GLCMLKLYQKRHPDINASAYSAYACLAIVIFFSVLGVVFGKGNTAFWIVFSIIHIIATLL 668

Query:  643 LSTQIYYMGRFKIDLGIFRRAAMVFYTDCIQQCSRPLYMDRMVLLVVGNLVNWSFALFGL 702
            LSTQ+YYMGR+K+D GIFRR   V YTDCI+QCS PLY+DRMVLLV+GN++NWS A +GL
Sbjct:  669 LSTQLYYMGRWKLDSGIFRRILHVLYTDCIRQCSGPLYVDRMVLLVMGNVINWSLAAYGL 728

Query:  703 IYRPRDFASYMLGIFICNLLLYLAFYIIMKLRSSEKVLPVPLFCIVATAVMWAAALYFFF 762
            I RP DFASY+L I ICNLLLY AFYIIMKLRS E++  +PL CIV T+V+W  AL+FFF
Sbjct:  729 IMRPNDFASYLLAIGICNLLLYFAFYIIMKLRSGERIKLIPLLCIVCTSVVWGFALFFFF 788

Query:  763 QNLSSWEGTPAESREKNRECILLDFFDDHDIWHFLSATALFFSFLV 808
            Q LS+W+ TPAESRE NR+CILLDFFDDHDIWHFLS+ A+F SFLV
Sbjct:  789 QGLSTWQKTPAESREHNRDCILLDFFDDHDIWHFLSSIAMFGSFLV 834
```

FIG. 1

BLASTP ALIGNMENT OF SEQ ID NO: 4, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH PROTEIN OF CLONE CT748_2 SEQ ID NO: 49

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 4)
Sbjct: W57901 Protein of clone CT748_2. [Homo sapiens] (SEQ ID NO: 49)
Length = 479

Score = 2220 (786.5 bits), Expect = 9.7e-231, P = 9.7e-231
Identities = 431/445 (96%), Positives = 433/445 (97%)

Query:  388 GGPPGQSDTDSSVEESDFDTMPDIESDKNIIRTKMFLYLSDLSRKDRRIVSKKYKIYFWN 447
            G    Q+ T + VEESDFDTMPDIESDKNIIRTKMFLYLSDLSRKDRRIVSKKYKIYFWN
Sbjct:   36 GHRASQTQT-APVEESDFDTMPDIESDKNIIRTKMFLYLSDLSRKDRRIVSKKYKIYFWN 94

Query:  448 IITIAVFYALPVIQLVITYQTVVNVTGNQDICYYNFLCAHPLGVLSAFNNILSNLGHVLL 507
            IITIAVFYALPVIQLVITYQTVVNVTGNQDICYYNFLCAHPLGVLSAFNNILSNLGHVLL
Sbjct:   95 IITIAVFYALPVIQLVITYQTVVNVTGNQDICYYNFLCAHPLGVLSAFNNILSNLGHVLL 154

Query:  508 GFLFLLIVLRRDILHRRALEAKDIFAVEYGIPKHFGLFYAMGIALMMEGVLSACYHVCPN 567
            GFLFLLIVLRRDILHRRALEAKDIFAVEYGIPKHFGLFYAMGIALMMEGVLSACYHVCPN
Sbjct:  155 GFLFLLIVLRRDILHRRALEAKDIFAVEYGIPKHFGLFYAMGIALMMEGVLSACYHVCPN 214

Query:  568 YSNFQFDTSFMYMIAGLCMLKLYQTRHPDINASAYSAYASFAVVIMVTVLGVVFGKNDVW 627
            YSNFQFDTSFMYMIAGLCMLKLYQTRHPDINASAYSAYASFAVVIMVTVLGVVFGKNDVW
Sbjct:  215 YSNFQFDTSFMYMIAGLCMLKLYQTRHPDINASAYSAYASFAVVIMVTVLGVVFGKNDVW 274

Query:  628 FWVIFSAIHVLASLALSTQIYYMGRFKID-----LGIFRRAAMVFYTDCIQQCSRPLYMD 682
            FWVIFSAIHVLASLALSTQIYYMGRFKID     LGIFRRAAMVFYTDCIQQCSRPLYMD
Sbjct:  275 FWVIFSAIHVLASLALSTQIYYMGRFKIDVSDTDLGIFRRAAMVFYTDCIQQCSRPLYMD 334

Query:  683 RMVLLVVGNLVNWSFALFGLIYRPRDFASYMLGIFICNLLLYLAFYIIMKLRSSEKVLPV 742
            RMVLLVVGNLVNWSFALFGLIYRPRDFASYMLGIFICNLLLYLAFYIIMKLRSSEKVLPV
Sbjct:  335 RMVLLVVGNLVNWSFALFGLIYRPRDFASYMLGIFICNLLLYLAFYIIMKLRSSEKVLPV 394

Query:  743 PLFCIVATAVMWAAALYFFFQNLSSWEGTPAESREKNRECILLDFFDDHDIWHFLSATAL 802
            PLFCIVATAVMWAAALYFFFQNLSSWEGTPAESREKNRECILLDFFDDHDIWHFLSATAL
Sbjct:  395 PLFCIVATAVMWAAALYFFFQNLSSWEGTPAESREKNRECILLDFFDDHDIWHFLSATAL 454

Query:  803 FFSFLVLLTLDDDLDVVRRDQIPVF 827
            FFSFLVLLTLDDDLDVVRRDQIPVF
Sbjct:  455 FFSFLVLLTLDDDLDVVRRDQIPVF 479
```

FIG. 2

BLASTP ALIGNMENT OF SEQ ID NO: 13, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH SIX TRANSMEMBRANE EPITHELIAL ANTIGEN OF PROSTATE SEQ ID NO: 50

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 13)
Sbjct: gi6572948 (AF186249) six transmembrane epithelial antigen of prostate [Homo sapiens] (SEQ ID NO: 50)
Length = 339

Score = 724 (259.9 bits), Expect = 2.3e-71, P = 2.3e-71
Identities = 126/267 (47%), Positives = 184/267 (68%)

Query:   208 LLPAWKVPTLLALGLFVCFYAYNFVRDVLQPYVQESQNKFFKLPVSVVNTTLPCVAYVLL 267
             L P W +P  +A  +    + Y  +R+V+ P      Q  F+K+P+ V+N  LP V+  LL
Sbjct:    67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Query:   268 SLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQHRKQIGLLSFFCAALHALYSFCLPLRR 327
             +LVYLPGV+AA +QL  GTKY++FP WLD W+  RKQ GLLSFF A LHA+YS   P+RR
Sbjct:   127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Query:   328 AHRYDLVNLAVKQVLANKSHLWVEEEVWRMEIYLSLGVLALGTLSLLAVTSLPSIANSLN 387
             ++RY L+N A +QV  NK   W+E +VWRMEIY+SLG++ L  L+LLAVTS+PS+++SL
Sbjct:   187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Query:   388 WREFSFVQSSLGFVALVLSTLHTLTYGWTRAFEESRYKFYLPPTFTLTLLVPCVVILAKA 447
             WREF ++QS LG V+L+L T+H L + W +  +  ++ +Y PPTF + + +P VV++ K+
Sbjct:   247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLIFKS 306

Query:   448 LFLLPCISRRLARIRRGWERESTIKPT 474
             +  LPC+ +++ +IR GWE  + I   T
Sbjct:   307 ILFLPCLRKKILKIRHGWEDVTKINKT 333
```

FIG. 3

BLASTP ALIGNMENT OF SEQ ID NO: 13, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH HUMAN STRAP-1 PROTEIN SEQ ID NO: 51

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 13)
Sbjct: Y58194 Human STRAP-1 protein. [Homo sapiens] (SEQ ID NO: 51)
Length = 339

Score = 724 (259.9 bits), Expect = 3.3e-72, P = 3.3e-72
Identities = 126/267 (47%), Positives = 184/267 (68%)

Query:  208 LLPAWKVPTLLALGLFVCFYAYNFVRDVLQPYVQESQNKFFKLPVSVVNTTLPCVAYVLL 267
            L P W +P  +A  +     + Y  +R+V+ P      Q. F+K+P+ V+N  LP V+  LL
Sbjct:   67 LFPQWHLPIKIAAIIASLTFLYTLLREVIHPLATSHQQYFYKIPILVINKVLPMVSITLL 126

Query:  268 SLVYLPGVLAAALQLRRGTKYQRFPDWLDHWLQHRKQIGLLSFFCAALHALYSFCLPLRR 327
            +LVYLPGV+AA +QL  GTKY++FP WLD W+  RKQ GLLSFF A LHA+YS  P+RR
Sbjct:  127 ALVYLPGVIAAIVQLHNGTKYKKFPHWLDKWMLTRKQFGLLSFFFAVLHAIYSLSYPMRR 186

Query:  328 AHRYDLVNLAVKQVLANKSHLWVEEEVWRMEIYLSLGVLALGTLSLLAVTSLPSIANSLN 387
            ++RY L+N A +QV  NK   W+E +VWRMEIY+SLG++ L  L+LLAVTS+PS+++SL
Sbjct:  187 SYRYKLLNWAYQQVQQNKEDAWIEHDVWRMEIYVSLGIVGLAILALLAVTSIPSVSDSLT 246

Query:  388 WREFSFVQSSLGFVALVLSTLHTLTYGWTRAFEESRYKFYLPPTFTLTLLVPCVVILAKA 447
            WREF ++QS LG V+L+L T+H L + W +  +  ++ +Y PPTF + + +P VV++ K+
Sbjct:  247 WREFHYIQSKLGIVSLLLGTIHALIFAWNKWIDIKQFVWYTPPTFMIAVFLPIVVLIFKS 306

Query:  448 LFLLPCISRRLARIRRGWERESTIKFT 474
            +  LPC+ +++ +IR GWE  + I  T
Sbjct:  307 ILFLPCLRKKILKIRHGWEDVTKINKT 333
```

FIG. 4

BLASTP ALIGNMENT OF SEQ ID NO: 20, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH SEVEN TRANSMEMBRANE RECEPTOR PROTEIN SEQ ID NO: 52

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 20)
Sbjct: gi5525078 (AB019120) seven transmembrane receptor [Rattus norvegicus] (SEQ ID NO: 52)
Length = 1349

Score = 5072 (1790.5 bits), Expect = 0.0, P = 0.0
Identities = 984/1354 (72%), Positives = 1108/1354 (81%)

Query:    1 MKSPRRTTLCLMFIVIYSSKAALNWNYESTIHPLSLHEHEPAGEEALRQKRAVATKSPTA 60
            MKS R  TL  + IVI SS+A  +    E  +HPL L EHE AGEE LR KRAVA    P A
Sbjct:    1 MKSSRTVTLYFVLIVICSSEATWSRPAEPIVHPLILQEHELAGEELLRPKRAVAVGGPVA 60

Query:   61 EEYTVNIEISFENASFLDPIKAYLNSLSFPIHGNNTDQITDILSINVTTVCRPAGNEIWC 120
            EEYTV++EISFEN SFL+ I+A+LNSL FP+ GN TD    ILS+ +TTVC P GN++ C
Sbjct:   61 EEYTVDVEISFENVSFLESIRAHLNSLRFPVQGNGTD----ILSMAMTTVCTPTGNDLLC 116

Query:  121 SCETGYGWPRERCLHNLICQERDVFLPGHHCSCLKELPPNGPFCLLQED-VTLNMRVRLN 179
              CE GY WP ERCL +L CQE D  LPG +C+CLK LPP GPFC L E  +TL ++VRLN
Sbjct:  117 FCEKGYQWPEERCLSSLTCQEHDSALPGRYCNCLKGLPPQGPFCQLPETYITLKIKVRLN 176

Query:  180 VGFQEDLMNTSSALYRSYKTDLETAFRKGYGILPGFKGVTVTGFKSGSVVVTYEVKTTPP 239
            +GFQEDL NTSSALYRSYKTDLE AFR GY  LPGF+ VTVT F  GSVVV Y V+
Sbjct:  177 IGFQEDLENTSSALYRSYKTDLERAFRAGYRTLPGFRSVTVTQFTKGSVVVDYIVEVASA 236

Query:  240 SLE-LIHKANEQVVQSLNQTYKMDYNSFQAVTINESNFFVTPEIIFEGDTVSLVCEKEVL 298
              L   IHKANEQV+Q+LNQTYKMDYNSFQ    NE+ F VTPE IFEGD V+L CE E +
Sbjct:  237 PLPGSIHKANEQVIQNLNQTYKMDYNSFQGTPSNETKFTVTPEFIFEGDNVTLECESEFV 296

Query:  299 SSNVSWRYEEQQLEIQNSSRFSIYTALFNNMTSVSKLTIHNITPGDAGEYVCKLILDIFE 358
            SSN SW Y E++ +IQNS +FSI+T++ NN++ V++LTI N T  DAG Y C + LDIFE
Sbjct:  297 SSNTSWFYGEKRSDIQNSDKFSIHTSIINNISLVTRLTIFNFTQHDAGLYGCNVTLDIFE 356
```

FIG. 5A

BLASTP ALIGNMENT OF SEQ ID NO: 20, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH SEVEN TRANSMEMBRANE RECEPTOR PROTEIN SEQ ID NO: 52

```
Query:  359 YECKKKIDVMPIQILANEEMKVMCDNNPVSLNCCSQGNVNWSKVEWKQEGKINIPGTPET 418
            Y   +K+DV PI+ILA EE KV+CDNNP+SLNCCS+   NWS++EWKQEGKINI GTPET
Sbjct:  357 YGTVRKLDVTPIRILAKEERKVVCDNNPISLNCCSENIANWSRIEWKQEGKINIEGTPET 416

Query:  419 DIDSSCSRYTLKADGTQCPSGSSGTTVIYTCEFISAYGARGSANIKVTFISVANLTITPD 478
            D++SSCS YTLKADGTQCPSGSSGTTVIYTCEF+S YGA+GS NI VTF SVANLTITPD
Sbjct:  417 DLESSCSTYTLKADGTQCPSGSSGTTVIYTCEFVSVYGAKGSKNIAVTFTSVANLTITPD 476

Query:  479 PISVSEGQNFSIKCISDVSNYDEVYWNTSAGIKIYQRFYTTRRYLDGAESVLTVKTSTRE 538
            PISVSEGQ+FSI C+SDVS++DEVYWNTSAGIKI+ RFYT RRY DGAESVLTVKTSTRE
Sbjct:  477 PISVSEGQSFSITCLSDVSSFDEVYWNTSAGIKIHPRFYTMRRYRDGAESVLTVKTSTRE 536

Query:  539 WNGTYHCIFRYKNSYSIATKDVIVHPLPLKLNIMVDPLEATVSCSGSHHIKCCIEE-DGD 597
            WNGTYHCIFRYKNSYSIATKDV VHPLPL+ +IM+DPLEA+  C+ SH  KCCIEE DG+
Sbjct:  537 WNGTYHCIFRYKNSYSIATKDVTVHPLPLESDIMMDPLEASGLCTSSHQFKCCIEENDGE 596

Query:  598 -YKVTFHMGSSSLPAAKEVNKKQVCYKHNFNASSVSWCSKTVDVCCHFTNAANNSVWSPS 656
             Y VTFH+ SSS PA +EV  KQ CY ++        S C K +DV CHFTNAAN+SV SPS
Sbjct:  597 EYIVTFHVDSSSPPAEREVIGKQACYTYSLPGKLPSRCPKDIDVFCHFTNAANSSVRSPS 656

Query:  657 MKLNLVPGENITCQDPVIGVGEPGKVIQKLCRFSNVPSSPESPIGGTITYKCVGSQWEEK 716
            MKL LVPG+NITCQDP+IG+GEPGKVIQKLC+F+ V  SP    IGGT+TYKCVGSQW+E+
Sbjct:  657 MKLTLVPGKNITCQDPIIGIGEPGKVIQKLCQFAGVSRSPGQTIGGTVTYKCVGSQWKEE 716

Query:  717 RNDCISAPINSLLQMAKALIKSPSQDEMLPTYLKDLSISIDKAEHEISSSPGSLGAIINI 776
               CISAPIN LLQ+AKALIKSPSQD+ LP YL+DLS+S  K E +I SSPGSLGAII+I
Sbjct:  717 TRACISAPINGLLQLAKALIKSPSQDQKLPKYLRDLSVSTGKEEQDIRSSPGSLGAIISI 776

Query:  777 LDLLSTVPTQVNSEMMTHVLSTVNVILGKPVLNTWKVLQQQWTNQSSQLLHSVERFSQAL 836
            LDLLSTVPTQVNSEMM  +L+T+NVIL K  LN+W+ L QQ +NQSSQ L SVERFS+AL
Sbjct:  777 LDLLSTVPTQVNSEMMRDILATINVILDKSTLNSWEKLLQQQSNQSSQFLQSVERFSKAL 836

Query:  837 QSGDS-PPLSFSQTNVQMSSTVIKSSHPETYQQRFVFPPYFDLWGNVVIDKSYLENLQSDS 895
            + GDS PP  F   NVQM S VIK  H + YQQ+FVF   DLWG+V ID+  L +LQ DS
Sbjct:  837 ELGDSTPPFLF-HPNVQMKSMVIKRGHAQMYQQKFVFTDSDLWGDVAIDECQLGSLQPDS 895
```

FIG 5B

BLASTP ALIGNMENT OF SEQ ID NO: 20, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH SEVEN TRANSMEMBRANE RECEPTOR PROTEIN SEQ ID NO: 52

```
Query:   896 SIVTMAFPTLQAILAQDIQENNFAESLVMTTTVSHNTTMPFRISMTFKNNSPSGGETKCV 955
             SIVT+AFPTL+AILAQD Q     + SLVMTTTVSHN    PFRISMTFKNN  SGG+ +CV
Sbjct:   896 SIVTVAFPTLKAILAQDGQRKTPSNSLVMTTTVSHNIVKPFRISMTFKNNHRSGGKPQCV 955

Query:   956 FWNFRLANNTGGWDSSGCYVEEGDGDN---VTCICDHLTSFSILMSPDSPDPSSLLGILL 1012
             FWNF LANNTGGWDSSGC VE+    DN     V C C+HLTSFSILMSPDSPDP SLL ILL
Sbjct:   956 FWNFSLANNTGGWDSSGCTVEDDGRDNRDRVFCKCNHLTSFSILMSPDSPDPGSLLKILL 1015

Query:  1013 DIISYVGVGFSILSLAACLVVEAVVWKSVTKNRTSYMRHTCIVNIAASLLVANTWFIVVA 1072
             DIISY+G+GFSI+SLAACLVVEA+VWKSVTKNRTSYMRH CIVNIA  LL+A+ WFIV
Sbjct:  1016 DIISYIGLGFSIVSLAACLVVEAMVWKSVTKNRTSYMRHICIVNIALCLLIADIWFIVAG 1075

Query:  1073 AIQDNRYILCKTACVAATFFIHFFYLSVFFWMLTLGLMLFYRLVFILHETSRSTQKAIAF 1132
             AI D   Y L +TACVAATFFIHFFYLSVFFWMLTLGLMLFYRL+FILH+ S+STQKAIAF
Sbjct:  1076 AIHDGHYPLNETACVAATFFIHFFYLSVFFWMLTLGLMLFYRLIFILHDASKSTQKAIAF 1135

Query:  1133 CLGYGCPLAISVITLGATQPREVYTRKNVCWLNWEDTKALLAFAIPALIIVVVNITITIV 1192
              LGYGCPL IS IT+G TQP+EVY RKN CWLNWEDT+ALLAFAIPALIIVVVN++IT+V
Sbjct:  1136 SLGYGCPLIISSITVGVTQPQEVYMRKNACWLNWEDTRALLAFAIPALIIVVVNVSITVV 1195

Query:  1193 VITKILRPSIGDKPCKQEKSSLFQISKSIGVLTPLLGLTWGFGLTTVFPGTNLVFHIIFA 1252
             VITKILRPS+GDKP KQEKSSLFQISKSIGVLTPLLGLTWGFGL TV  G+N VFHIIF
Sbjct:  1196 VITKILRPSVGDKPGKQEKSSLFQISKSIGVLTPLLGLTWGFGLATVIQGSNAVFHIIFT 1255

Query:  1253 ILNVFQGLFILLFGCLWDLKVQEALLNKFSLSRWSSQHSKSTSLGSSTPVFSMSSPISRR 1312
             +LN FQGLFILLFGCLWD KVQEALL+KFSLSRWSSQHSKSTSLGSSTPVFSMSSPISRR
Sbjct:  1256 LLNAFQGLFILLFGCLWDQKVQEALLHKFSLSRWSSQHSKSTSLGSSTPVFSMSSPISRR 1315

Query:  1313 FNNLFGKTGTYNVSTPEATSSSLENSSSASSLLN 1346
             FNNLFGKTGTYNVSTPE TSSS+ENSSSA SLLN
Sbjct:  1316 FNNLFGKTGTYNVSTPETTSSSVENSSSAYSLLN 1349
```

FIG. 5C

BLASTP ALIGNMENT OF SEQ ID NO: 20, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH HUMAN BRAIN-DERIVED G PROTEIN-COUPLED RECEPTOR PROTEIN SEQ ID NO: 53

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 20)
Sbjct: Y40440 Human brain-derived G-protein coupled receptor protein. [Homo sapiens] (SEQ ID NO: 53)
Length = 986

Score = 5161 (1821.8 bits), Expect = 0.0, P = 0.0
Identities = 986/986 (100%), Positives = 986/986 (100%)

Query:   361 CKKKIDVMPIQILANEEMKVMCDNNPVSLNCCSQGNVNWSKVEWKQEGKINIPGTPETDI 420
             CKKKIDVMPIQILANEEMKVMCDNNPVSLNCCSQGNVNWSKVEWKQEGKINIPGTPETDI
Sbjct:     1 CKKKIDVMPIQILANEEMKVMCDNNPVSLNCCSQGNVNWSKVEWKQEGKINIPGTPETDI 60

Query:   421 DSSCSRYTLKADGTQCPSGSSGTTVIYTCEFISAYGARGSANIKVTFISVANLTITPDPI 480
             DSSCSRYTLKADGTQCPSGSSGTTVIYTCEFISAYGARGSANIKVTFISVANLTITPDPI
Sbjct:    61 DSSCSRYTLKADGTQCPSGSSGTTVIYTCEFISAYGARGSANIKVTFISVANLTITPDPI 120

Query:   481 SVSEGQNFSIKCISDVSNYDEVYWNTSAGIKIYQRFYTTRRYLDGAESVLTVKTSTREWN 540
             SVSEGQNFSIKCISDVSNYDEVYWNTSAGIKIYQRFYTTRRYLDGAESVLTVKTSTREWN
Sbjct:   121 SVSEGQNFSIKCISDVSNYDEVYWNTSAGIKIYQRFYTTRRYLDGAESVLTVKTSTREWN 180

Query:   541 GTYHCIFRYKNSYSIATKDVIVHPLPLKLNIMVDPLEATVSCSGSHHIKCCIEEDGDYKV 600
             GTYHCIFRYKNSYSIATKDVIVHPLPLKLNIMVDPLEATVSCSGSHHIKCCIEEDGDYKV
Sbjct:   181 GTYHCIFRYKNSYSIATKDVIVHPLPLKLNIMVDPLEATVSCSGSHHIKCCIEEDGDYKV 240

Query:   601 TFHMGSSSLPAAKEVNKKQVCYKHNFNASSVSWCSKTVDVCCHFTNAANNSVWSPSMKLN 660
             TFHMGSSSLPAAKEVNKKQVCYKHNFNASSVSWCSKTVDVCCHFTNAANNSVWSPSMKLN
Sbjct:   241 TFHMGSSSLPAAKEVNKKQVCYKHNFNASSVSWCSKTVDVCCHFTNAANNSVWSPSMKLN 300

Query:   661 LVPGENITCQDPVIGVGEPGKVIQKLCRFSNVPSSPESPIGGTITYKCVGSQWEEKRNDC 720
             LVPGENITCQDPVIGVGEPGKVIQKLCRFSNVPSSPESPIGGTITYKCVGSQWEEKRNDC
Sbjct:   301 LVPGENITCQDPVIGVGEPGKVIQKLCRFSNVPSSPESPIGGTITYKCVGSQWEEKRNDC 360

Query:   721 ISAPINSLLQMAKALIKSPSQDEMLPTYLKDLSISIDKAEHEISSSPGSLGAIINILDLL 780
             ISAPINSLLQMAKALIKSPSQDEMLPTYLKDLSISIDKAEHEISSSPGSLGAIINILDLL
Sbjct:   361 ISAPINSLLQMAKALIKSPSQDEMLPTYLKDLSISIDKAEHEISSSPGSLGAIINILDLL 420

Query:   781 STVPTQVNSEMMTHVLSTVNVILGKPVLNTWKVLQQQWTNQSSQLLHSVERFSQALQSGD 840
             STVPTQVNSEMMTHVLSTVNVILGKPVLNTWKVLQQQWTNQSSQLLHSVERFSQALQSGD
Sbjct:   421 STVPTQVNSEMMTHVLSTVNVILGKPVLNTWKVLQQQWTNQSSQLLHSVERFSQALQSGD 480
```

FIG. 6A

BLASTP ALIGNMENT OF SEQ ID NO: 20, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH HUMAN BRAIN-DERIVED G PROTEIN-COUPLED RECEPTOR PROTEIN SEQ ID NO: 53

```
Query:   841 SPPLSFSQTNVQMSSTVIKSSHPETYQQRFVFPYFDLWGNVVIDKSYLENLQSDSSIVTM 900
             SPPLSFSQTNVQMSSTVIKSSHPETYQQRFVFPYFDLWGNVVIDKSYLENLQSDSSIVTM
Sbjct:   481 SPPLSFSQTNVQMSSTVIKSSHPETYQQRFVFPYFDLWGNVVIDKSYLENLQSDSSIVTM 540

Query:   901 AFPTLQAILAQDIQENNFAESLVMTTTVSHNTTMPFRISMTFKNNSPSGGETKCVFWNFR 960
             AFPTLQAILAQDIQENNFAESLVMTTTVSHNTTMPFRISMTFKNNSPSGGETKCVFWNFR
Sbjct:   541 AFPTLQAILAQDIQENNFAESLVMTTTVSHNTTMPFRISMTFKNNSPSGGETKCVFWNFR 600

Query:   961 LANNTGGWDSSGCYVEEGDGDNVTCICDHLTSFSILMSPDSPDPSSLLGILLDIISYVGV 1020
             LANNTGGWDSSGCYVEEGDGDNVTCICDHLTSFSILMSPDSPDPSSLLGILLDIISYVGV
Sbjct:   601 LANNTGGWDSSGCYVEEGDGDNVTCICDHLTSFSILMSPDSPDPSSLLGILLDIISYVGV 660

Query:  1021 GFSILSLAACLVVEAVVWKSVTKNRTSYMRHTCIVNIAASLLVANTWFIVVAAIQDNRYI 1080
             GFSILSLAACLVVEAVVWKSVTKNRTSYMRHTCIVNIAASLLVANTWFIVVAAIQDNRYI
Sbjct:   661 GFSILSLAACLVVEAVVWKSVTKNRTSYMRHTCIVNIAASLLVANTWFIVVAAIQDNRYI 720

Query:  1081 LCKTACVAATFFIHFFYLSVFFWMLTLGLMLFYRLVFILHETSRSTQKAIAFCLGYGCPL 1140
             LCKTACVAATFFIHFFYLSVFFWMLTLGLMLFYRLVFILHETSRSTQKAIAFCLGYGCPL
Sbjct:   721 LCKTACVAATFFIHFFYLSVFFWMLTLGLMLFYRLVFILHETSRSTQKAIAFCLGYGCPL 780

Query:  1141 AISVITLGATQPREVYTRKNVCWLNWEDTKALLAFAIPALIIVVVNITITIVVITKILRP 1200
             AISVITLGATQPREVYTRKNVCWLNWEDTKALLAFAIPALIIVVVNITITIVVITKILRP
Sbjct:   781 AISVITLGATQPREVYTRKNVCWLNWEDTKALLAFAIPALIIVVVNITITIVVITKILRP 840

Query:  1201 SIGDKPCKQEKSSLFQISKSIGVLTPLLGLTWGFGLTTVFPGTNLVFHIIFAILNVFQGL 1260
             SIGDKPCKQEKSSLFQISKSIGVLTPLLGLTWGFGLTTVFPGTNLVFHIIFAILNVFQGL
Sbjct:   841 SIGDKPCKQEKSSLFQISKSIGVLTPLLGLTWGFGLTTVFPGTNLVFHIIFAILNVFQGL 900

Query:  1261 FILLFGCLWDLKVQEALLNKFSLSRWSSQHSKSTSLGSSTPVFSMSSPISRRFNNLFGKT 1320
             FILLFGCLWDLKVQEALLNKFSLSRWSSQHSKSTSLGSSTPVFSMSSPISRRFNNLFGKT
Sbjct:   901 FILLFGCLWDLKVQEALLNKFSLSRWSSQHSKSTSLGSSTPVFSMSSPISRRFNNLFGKT 960

Query:  1321 GTYNVSTPEATSSSLENSSSASSLLN 1346
             GTYNVSTPEATSSSLENSSSASSLLN
Sbjct:   961 GTYNVSTPEATSSSLENSSSASSLLN 986
```

FIG. 6B

BLASTP ALIGNMENT OF SEQ ID NO: 29, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH PUTATIVE SEVEN PASS TRANSMEMBRANE PROTEIN SEQ ID NO: 54

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 29)
Sbjct: gi8132351 (AF154337) putative seven pass transmembrane protein [Mus musculus] (SEQ ID NO: 54)
Length = 385

Score = 946 (338.1 bits), Expect = 6.7e-95, P = 6.7e-95
Identities = 186/323 (57%), Positives = 234/323 (72%)

Query:    14 LVPALPPAVTLGLTAAYTTLYALLFFSVYAQLWLVLLYGHKRLSYQTVFLALCLLWAALR 73
             L PA+PP V LGLTA YT  YALLF  +YAQLWLVL Y HKRLSYQ+VFL LCL WA+LR
Sbjct:    18 LSPAVPPYVKLGLTAVYTVFYALLFVFIYAQLWLVLRYRHKRLSYQSVFLFLCLFWASLR 77

Query:    74 TTLFSFYFRDTPRANRLGPLPFWLLYCCPVCLQFFTLTLMNLYFAQVVFKAKVKRRPEMS 133
             T LFSFYFRD   AN    P  FWLLYC PVCLQFFTLTLMNLYF QV+FKAK K  PE+
Sbjct:    78 TVLFSFYFRDFVAANSFSPFVFWLLYCFPVCLQFFTLTLMNLYFTQVIFKAKSKYSPELL 137

Query:   134 RGLLAVRGAFVGASLLFLLVNVLCAVLSHRRRAQPWALLLVRVLVSDSLFVICALSLAAC 193
             +  L +  A +  SL+FLLVN+ CAVL          ++ VRV ++D+LFV+CA+SL+ C
Sbjct:   138 KYRLPLYLASLFISLVFLLVNLTCAVLVKTGDWDRKVIVSVRVAINDTLFVLCAISLSIC 197

Query:   194 LCLVARRAPSTSIYLEAKGTSVCQAAAMGGAMVLLYASRACYNLTALALAPQSRLDTFDY 253
             L  +++ + + +IYLE+KG+SVCQ  A+G  ++LLYASRACYNL  L+ +     + +FDY
Sbjct:   198 LYKISKMSLA-NIYLESKGSSVCQVTAIGVTVILLYASRACYNLFILSFSQIKNVHSFDY 256

Query:   254 DWYNVSDQADLVNDLGNKGYLVFGLILFVWELLPTTLLVGFFRVHRPPQDLSTSHILNGQ 313
             DWYNVSDQADL + LG+ GY+VFG++LFVWELLPTTL+V FFRV  P +DL+   ++
Sbjct:   257 DWYNVSDQADLKSQLGDAGYVVFGVVLFVWELLPTTLVVYFFRVRNPTKDLTNPGMVPSH 316

Query:   314 VFASRSYFFDRAGHCE-DEGCSW 335
             F+ RSYFFD     + D+  +W
Sbjct:   317 GFSPRSYFFDNPRRYDSDDDLAW 339
```

FIG. 7

BLASTP ALIGNMENT OF SEQ ID NO: 29 G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH HUMAN h-TRAAK POLYPEPTIDE #1 SEQ ID NO: 55

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 29)
Sbjct: Y94425 Human h-TRAAK polypeptide #1  [Homo sapiens] (SEQ ID NO: 55)
Length = 393

Score = 2062 (730.9 bits), Expect = 5.4e-214, P = 5.4e-214
Identities = 392/392 (100%), Positives = 392/392 (100%)

Query:   540 MRSTTLLALLALVLLYLVSGALVFRALEQPHEQQAQRELGEVREKFLRAHPCVSDQELGL 599
             MRSTTLLALLALVLLYLVSGALVFRALEQPHEQQAQRELGEVREKFLRAHPCVSDQELGL
Sbjct:     1 MRSTTLLALLALVLLYLVSGALVFRALEQPHEQQAQRELGEVREKFLRAHPCVSDQELGL 60

Query:   600 LIKEVADALGGGADPETNSTSNSSHSAWDLGSAFFFSGTIITTIGYGNVALRTDAGRLFC 659
             LIKEVADALGGGADPETNSTSNSSHSAWDLGSAFFFSGTIITTIGYGNVALRTDAGRLFC
Sbjct:    61 LIKEVADALGGGADPETNSTSNSSHSAWDLGSAFFFSGTIITTIGYGNVALRTDAGRLFC 120

Query:   660 IFYALVGIPLFGILLAGVGDRLGSSLRHGIGHIEAIFLKWHVPPELVRVLSAMLFLLIGC 719
             IFYALVGIPLFGILLAGVGDRLGSSLRHGIGHIEAIFLKWHVPPELVRVLSAMLFLLIGC
Sbjct:   121 IFYALVGIPLFGILLAGVGDRLGSSLRHGIGHIEAIFLKWHVPPELVRVLSAMLFLLIGC 180

Query:   720 LLFVLTPTFVFCYMEDWSKLEAIYFVIVTLTTVGFGDYVAGADPRQDSPAYQPLVWFWIL 779
             LLFVLTPTFVFCYMEDWSKLEAIYFVIVTLTTVGFGDYVAGADPRQDSPAYQPLVWFWIL
Sbjct:   181 LLFVLTPTFVFCYMEDWSKLEAIYFVIVTLTTVGFGDYVAGADPRQDSPAYQPLVWFWIL 240

Query:   780 LGLAYFASVLTTIGNWLRVVSRRTRAEMGGLTAQAASWTGTVTARVTQRAGPAAPPPEKE 839
             LGLAYFASVLTTIGNWLRVVSRRTRAEMGGLTAQAASWTGTVTARVTQRAGPAAPPPEKE
Sbjct:   241 LGLAYFASVLTTIGNWLRVVSRRTRAEMGGLTAQAASWTGTVTARVTQRAGPAAPPPEKE 300

Query:   840 QPLLPPPPCPAQPLGRPRSPSPPEKAQPPSPPTASALDYPSENLAFIDESSDTQSERGCP 899
             QPLLPPPPCPAQPLGRPRSPSPPEKAQPPSPPTASALDYPSENLAFIDESSDTQSERGCP
Sbjct:   301 QPLLPPPPCPAQPLGRPRSPSPPEKAQPPSPPTASALDYPSENLAFIDESSDTQSERGCP 360

Query:   900 LPRAPRGRRRPNPPRKPVRPRGPGRPRDKGVP 931
             LPRAPRGRRRPNPPRKPVRPRGPGRPRDKGVP
Sbjct:   361 LPRAPRGRRRPNPPRKPVRPRGPGRPRDKGVP 392
```

FIG. 8

BLASTP ALIGNMENT OF SEQ ID NO: 36, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH HUMAN OLFACTORY RECEPTOR POLYPEPTIDE SEQ ID NO: 56

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 36)
Sbjct: gi2921716 (U86281) olfactory receptor [Homo sapiens] (SEQ ID NO: 56)
Length = 217

Score = 779 (279.3 bits), Expect = 3.3e-77, P = 3.3e-77
Identities = 146/166 (87%), Positives = 154/166 (92%)

Query:     1 MSYDRYVAICHPLRYFIIMTWKVCITLAITSWTCGSLLAMVHVSLILRLPFCGPREINHF 60
             MSYDRY+AICHPL+Y +IM W VC  LA+TSW CGSLLA+VHV LILRLPFCGP EINHF
Sbjct:    52 MSYDRYMAICHPLQYSVIMRWGVCTVLAVTSWACGSLLALVHVVLILRLPFCGPHEINHF 111

Query:    61 FCEILSVLRLACADTWLNQVVIFAACMFILVGPLCLVLVSYSHILAAILRIQSGEGRRKA 120
             FCEILSVL+LACADTWLNQVVIFAA +FILVGPLCLVLVSYS ILAAILRIQSGEGRRKA
Sbjct:   112 FCEILSVLKLACADTWLNQVVIFAASVFILVGPLCLVLVSYSRILAAILRIQSGEGRRKA 171

Query:   121 FSTCSSHLCVVGLFFGSAIVMYMAPKSRHPEEQQKVLFLFYSSFNP 166
             FSTCSSHLC+VGLFFGSAIVMYMAPKSRHPEEQQKVL LFYS FNP
Sbjct:   172 FSTCSSHLCMVGLFFGSAIVMYMAPKSRHPEEQQKVLSLFYSLFNP 217
```

FIG. 9

BLASTP ALIGNMENT OF SEQ ID NO: 36 G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH HUMAN G PROTEIN-COUPLED RECEPTOR GPR1 POLYPEPTIDE SEQ ID NO: 57

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 36)
Sbjct: W04244 Human G-protein coupled receptor GPR1 [Homo sapiens] (SEQ ID NO: 57)
Length = 296

 Score = 835 (299.0 bits), Expect = 5.7e-84, P = 5.7e-84
 Identities = 159/171 (92%), Positives = 160/171 (93%)

Query:    1 MSYDRYVAICHPLRYFIIMTWKVCITLAITSWTCGSLLAMVHVSLILRLPFCGPREINHF 60
            MSYDRYVAICHPLRYFIIMTWKVCITL ITSWTCGSLLAMVHVSLILRLPFCGPREINHF
Sbjct:  121 MSYDRYVAICHPLRYFIIMTWKVCITLGITSWTCGSLLAMVHVSLILRLPFCGPREINHF 180

Query:   61 FCEILSVLRLACADTWLNQVVIFAACMFILVGPLCLVLVSYSHILAAILRIQSGEGRRKA 120
            FCEILSVLRLACADTWLNQVVIF ACMFILVGPLCLVLVSYSHIL  ILRIQSGEGRRKA
Sbjct:  181 FCEILSVLRLACADTWLNQVVIFEACMFILVGPLCLVLVSYSHILGGILRIQSGEGRRKA 240

Query:  121 FSTCSSHLCVVGLFFGSAIVMYMAPKSRHPEEQQKVLFLFYSSFN-PMLNP 170
            FSTCSSHLCVVGLFFGSAIVMYMAPKSRHPEEQQKVLFL      + PML P
Sbjct:  241 FSTCSSHLCVVGLFFGSAIVMYMAPKSRHPEEQQKVLFLILQFLSTPMLKP 291
```

FIG. 10

BLASTP ALIGNMENT OF SEQ ID NO: 42, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH SIMILAR TO MOUSE OLFACTORY RECEPTOR 13 POLYPEPTIDE SEQ ID NO: 58

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 42)
Sbjct: gi4159884 (AC005587) similar to mouse olfactory receptor 13;
 similar to P34984 (PID:g464305) [Homo sapiens] (SEQ ID NO: 58)
Length = 310

Score = 1067 (380.7 bits), Expect = 1.0e-107, P = 1.0e-107
Identities = 209/304 (68%), Positives = 247/304 (81%)

Query:    1 MVKNQTMVTEFLLLGFLLGPRIQMLLFGLFSLFYIFTLLGNGAILGLISLDSRLHTPMYF 60
            M  N T + EFLLLGF +GPRIQMLLFGLFSLFY+FTLLGNG ILGLISLDSRLH PMYF
Sbjct:    1 MGDNITSIREFLLLGFPVGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHAPMYF 60

Query:   61 FLSHLAVVDIAYTRNTVPQMLANLLHPAKPISFAGCMTQTFLCLSFGHSECLLLVLMSYD 120
            FLSHLAVVDIAY  NTVP+ML NLLHPAKPISFAG M QTFL  +F  +ECLLLV+MSYD
Sbjct:   61 FLSHLAVVDIAYACNTVPRMLVNLLHPAKPISFAGRMMQTFLFSTFAVTECLLLVVMSYD 120

Query:  121 RYVAICHPLRYSVIMT--CCITLAITSWTCGSLLAMVHVSLILRLPFCGPREINHFFCEI 178
             YVAICHPLRY  IMT  CITLA+TSWT G LL+++H+ L+L LPFC P++I HFFCEI
Sbjct:  121 LYVAICHPLRYLAIMTWRVCITLAVTSWTTGVLLSLIHLVLLLPLPFCRPQKIYHFFCEI 180

Query:  179 LSVLRLACADTWLNQVVIFAACMFILVGPLCLVLVSYSHILAAILRIQSGEGRRKAFSTC 238
            L+VL+LACADT +N+ ++ A  +  LVGPL  ++VSY  IL AIL+IQS E +RKAF TC
Sbjct:  181 LAVLKLACADTHINENMVLAGAISGLVGPLSTIVVSYMCILCAILQIQSREVQRKAFRTC 240

Query:  239 SSHLCVVGLFFGSAIVMYMAPKSRHPEEQQKVLFLFYSSFNPMLNPLIYNLRNVEVKGAL 298
             SHLCV+GL +G+AI+MY+ P+  +P+EQ+K L LF+S FNPMLNPLI +LRN EVK  L
Sbjct:  241 FSHLCVIGLVYGTAIIMYVGPRYGNPKEQKKYLLLFHSLFNPMLNPLICSLRNSEVKNTL 300

Query:  299 RRAL 302
            +R L
Sbjct:  301 KRVL 304
```

FIG. 11

BLASTP ALIGNMENT OF SEQ ID NO: 42, G PROTEIN-COUPLED RECEPTOR-LIKE POLYPEPTIDE (IDENTIFIED AS GPCR-LIKE) WITH HUMAN G PROTEIN-COUPLED RECEPTOR GPR1 POLYPEPTIDE SEQ ID NO: 59

```
Query: G protein-coupled receptor-like polypeptide (SEQ ID NO: 42)
Sbjct: W04244 Human G-protein coupled receptor GPR1. [Homo sapiens] (SEQ ID NO: 59)
Length = 296

Score = 1328 (472.5 bits), Expect = 3.2e-136, P = 3.2e-136
Identities = 259/287 (90%), Positives = 264/287 (91%)

Query:    1 MVKNQTMVTEFLLLGFLLGPRIQMLLFGLFSLFYIFTLLGNGAILGLISLDSRLHTPMYF 60
            MVKNQTMVTEFLLLGFLLGPRIQMLLFGLFSLFY+FTLLGNG ILGLISLDSRLHTPMYF
Sbjct:    5 MVKNQTMVTEFLLLGFLLGPRIQMLLFGLFSLFYVFTLLGNGTILGLISLDSRLHTPMYF 64

Query:   61 FLSHLAVVDIAYTRNTVPQMLANLLHPAKPISFAGCMTQTFLCLSFGHSECLLLVLMSYD 120
            FLSHLAVV+IAY  NTVPQML NLLHPAKPISFAGCMT  FL LSF H+ECLLLVLMSYD
Sbjct:   65 FLSHLAVVNIAYACNTVPQMLVNLLHPAKPISFAGCMTLDFLFLSFAHTECLLLVLMSYD 124

Query:  121 RYVAICHPLRYSVIMT--CCITLAITSWTCGSLLAMVHVSLILRLPFCGPREINHFFCEI 178
            RYVAICHPLRY +IMT   CITL ITSWTCGSLLAMVHVSLILRLPFCGPREINHFFCEI
Sbjct:  125 RYVAICHPLRYFIIMTWKVCITLGITSWTCGSLLAMVHVSLILRLPFCGPREINHFFCEI 184

Query:  179 LSVLRLACADTWLNQVVIFAACMFILVGPLCLVLVSYSHILAAILRIQSGEGRRKAFSTC 238
            LSVLRLACADTWLNQVVIF ACMFILVGPLCLVLVSYSHIL  ILRIQSGEGRRKAFSTC
Sbjct:  185 LSVLRLACADTWLNQVVIFEACMFILVGPLCLVLVSYSHILGGILRIQSGEGRRKAFSTC 244

Query:  239 SSHLCVVGLFFGSAIVMYMAPKSRHPEEQQKVLFLFYSSFN-PMLNP 284
            SSHLCVVGLFFGSAIVMYMAPKSRHPEEQQKVLFL      + PML P
Sbjct:  245 SSHLCVVGLFFGSAIVMYMAPKSRHPEEQQKVLFLILQFLSTPMLKP 291
```

FIG. 12

SEQUENCE LISTING

<110> HYSEQ, Inc.
Yamazaki, Victoria
Tang, Y. Tom
Liu, Chenghua
Zhou, Ping
Wang, Dunrui
Zhang, Jie
Ren, Feiyan
Asundi, Vinod
Drmanac, Radoje T

<120> METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE (GPCR-LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES

<130> HYS-37CIP

<140> not yet assigned
<141> 2000-12-22

<150> not yet assigned
<151> 2000-12-21

<150> US 09/729,739
<151> 2000-12-04

<150> US 09/653,450
<151> 2000-08-31

<150> US 09/620,312
<151> 2000-07-19

<150> US 09/598,042
<151> 2000-06-20

<150> US 09/552,317
<151> 2000-04-25

<150> US 09/488,725
<151> 2000-01-21

<160> 63

<170> PatentIn version 3.0

<210> 1
<211> 412
<212> DNA
<213> Homo sapiens

<400> 1
cttttacatc atcatgaagc tccgcagctc tgaaaaggtc ctcccagtcc cgctcttctg 60

catcgtggcc accgctgtga tgtgggctgc cgccctatat tttttcttcc agaatctcag 120

cagctgggag ggaactccgg ccgaatcccg ggagaagaac cgcgagtgca ttctgctgga 180

tttcttcgat gaccatgaca tctggcactt cctctctgct actgctctgt ttttctcatt 240

cttggatttg ttaactttgg atgatgacct tgatgtggtt cggagagacc agatccctgt 300

cttctgaacc tccaacatta agagagggga gggagcgatc aatcttggtg ctgtttcaca 360

aaaattacag tgaccacagc aaagtaacca ctgccagatg ctccactcac cc 412

<210> 2
<211> 749
<212> DNA
<213> Homo sapiens

<400> 2
cttttacatc atcatgaagc tccgcagctc tgaaaaggtc ctcccagtcc cgctcttctg 60

catcgtggcc accgctgtga tgtgggctgc cgccctatat tttttcttcc agaatctcag 120

cagctgggag ggaactccgg ccgaatcccg ggagaagaac cgcgagtgca ttctgctgga 180

tttcttcgat gaccatgaca tctggcactt cctctctgct actgctctgt tttctcatt 240

cttggatttg ttaactttgg atgatgacct tgatgtggtt cggagagacc agatccctgt 300

cttctgaacc tccaacatta agagagggga gggagcgatc aatcttggtg ctgtttcaca 360

aaaattacag tgaccacagc aaagtaacca ctgccagatg ctccactcac cctctgtaga 420

gccaactctg cattcacaca ggaaggagag gggctgcggg agatttaaac ctgcaagaaa 480

ggaggcagaa ggggagccat gttttgagga cagacgcaaa cctgaggagc tgagaaacac 540

ttgctccttc catctgcagc tttgggagtg caacagggat aggcactgca tccaagtcaa 600

ctcaccatct tggggtccct cccaccctca cggagacttg ccagcaatgg cagaatgctg 660

ctgcacactt tccttcaagt gttacccttc ccagaaaggc caagctcgtg gacttcttgg 720

ccaaatactt gggtaggccc cgcgttccg 749

<210> 3
<211> 3188
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(2484)

<400> 3
atg cgc ggc tgc ctg cgg ctc gcg ctg ctc tgc gcg ctg ccc tgg ctc 48
Met Arg Gly Cys Leu Arg Leu Ala Leu Leu Cys Ala Leu Pro Trp Leu
1 5 10 15

ctg ctg gcg gcg tcg ccc ggg cac ccg gcg aaa tcc ccc agg cag ccc 96
Leu Leu Ala Ala Ser Pro Gly His Pro Ala Lys Ser Pro Arg Gln Pro
20 25 30

ccg gca ccg cgc cgc gac ccc ttc gac gct gcc agg ggc gcc gat ttc 144
Pro Ala Pro Arg Arg Asp Pro Phe Asp Ala Ala Arg Gly Ala Asp Phe
35 40 45

gat cat gtc tac agc ggg gtg gtg aac ctc agc acc gag aac atc tac 192
Asp His Val Tyr Ser Gly Val Val Asn Leu Ser Thr Glu Asn Ile Tyr
50 55 60

tct ttc aac tac acc agc cag ccc gac cag gtg aca gcc gtg agg gtg 240

Ser Phe Asn Tyr Thr Ser Gln Pro Asp Gln Val Thr Ala Val Arg Val
65 70 75 80

tat gtg aac agt tcc tct gag aat ctc aac tac ccg gtc ctt gtt gtg 288
Tyr Val Asn Ser Ser Ser Glu Asn Leu Asn Tyr Pro Val Leu Val Val
85 90 95

gtt cgc cag cag aaa gag gtg ctg tcc tgg cag gtt cct ctg ctc ttc 336
Val Arg Gln Gln Lys Glu Val Leu Ser Trp Gln Val Pro Leu Leu Phe
100 105 110

caa gga cta tac cag agg agc tac aat tat caa gaa gtg agc cgc acc 384
Gln Gly Leu Tyr Gln Arg Ser Tyr Asn Tyr Gln Glu Val Ser Arg Thr
115 120 125

tta tgt ccc tca gaa gca acc aat gag acg gga ccc ttg cag caa ctg 432
Leu Cys Pro Ser Glu Ala Thr Asn Glu Thr Gly Pro Leu Gln Gln Leu
130 135 140

ata ttt gta gat gtc gca tcc atg gca ccc ctg ggt gct cag tac aaa 480
Ile Phe Val Asp Val Ala Ser Met Ala Pro Leu Gly Ala Gln Tyr Lys
145 150 155 160

ctg cta gtt acc aag ctg aag cac ttc cag ctc cgg aca aat gtt gcc 528
Leu Leu Val Thr Lys Leu Lys His Phe Gln Leu Arg Thr Asn Val Ala
165 170 175

ttt cac ttt act gcc agc ccc tct caa cct cag tat ttt cta tac aag 576
Phe His Phe Thr Ala Ser Pro Ser Gln Pro Gln Tyr Phe Leu Tyr Lys
180 185 190

ttt ccc aaa gac gtg gac tca gtt atc att aaa gtg gtg tct gaa atg 624
Phe Pro Lys Asp Val Asp Ser Val Ile Ile Lys Val Val Ser Glu Met
195 200 205

gct tat cca tgt tct gtt gtc tca gtc cag aat atc atg tgc ccg gtg 672
Ala Tyr Pro Cys Ser Val Val Ser Val Gln Asn Ile Met Cys Pro Val
210 215 220

tat gat ctc gac cac aat gtg gaa ttt aat ggt gtc tat cag tcc atg 720
Tyr Asp Leu Asp His Asn Val Glu Phe Asn Gly Val Tyr Gln Ser Met
225 230 235 240

acc aag aaa gct gcc atc acg cta cag aag aag gat ttt cca ggc gag 768
Thr Lys Lys Ala Ala Ile Thr Leu Gln Lys Lys Asp Phe Pro Gly Glu
245 250 255

cag ttc ttc gtg gta ttt gtg ata aag cct gaa gat tat gcc tgt gga 816
Gln Phe Phe Val Val Phe Val Ile Lys Pro Glu Asp Tyr Ala Cys Gly
260 265 270

gga tct ttc ttc atc cag gaa aag gaa aac cag acc tgg aat cta cag 864
Gly Ser Phe Phe Ile Gln Glu Lys Glu Asn Gln Thr Trp Asn Leu Gln
275 280 285

cga aaa aag aac ctt gaa gtg acc att gtc cct tcc att aaa gaa tct 912
Arg Lys Lys Asn Leu Glu Val Thr Ile Val Pro Ser Ile Lys Glu Ser
290 295 300

gtt tat gtg aaa tcc agt ctt ttc agt gtc ttc atc ttc ctg tcc ttc 960
Val Tyr Val Lys Ser Ser Leu Phe Ser Val Phe Ile Phe Leu Ser Phe
305 310 315 320

tac ttg gga tgc ctt ctt gtt ggg ttt gtt cat tat ctg agg ttt cag 1008
Tyr Leu Gly Cys Leu Leu Val Gly Phe Val His Tyr Leu Arg Phe Gln
325 330 335

aga aaa tcc att gat gga agc ttt ggg tcc aat gat ggc tct gga aat 1056
Arg Lys Ser Ile Asp Gly Ser Phe Gly Ser Asn Asp Gly Ser Gly Asn
340 345 350

atg gtg gca tct cat ccc att gct gcc agc aca ccc gaa ggg agc aat 1104
Met Val Ala Ser His Pro Ile Ala Ala Ser Thr Pro Glu Gly Ser Asn
355 360 365

tat ggg aca ata gat gag tca agc tcc agt cct gga agg cag atg tcc 1152
Tyr Gly Thr Ile Asp Glu Ser Ser Ser Ser Pro Gly Arg Gln Met Ser
370 375 380

tcc tcc gat ggt ggg cca ccg ggc cag tca gac aca gac agc tcc gtg 1200
Ser Ser Asp Gly Gly Pro Pro Gly Gln Ser Asp Thr Asp Ser Ser Val
385 390 395 400

gag gag agc gac ttc gac acc atg cca gac att gag agt gat aaa aac 1248
Glu Glu Ser Asp Phe Asp Thr Met Pro Asp Ile Glu Ser Asp Lys Asn
405 410 415

atc atc cgg acc aag atg ttc ctt tac ctg tca gat ttg tcc agg aag 1296
Ile Ile Arg Thr Lys Met Phe Leu Tyr Leu Ser Asp Leu Ser Arg Lys
420 425 430

gac cgg aga att gtc agc aaa aaa tat aaa att tat ttt tgg aac atc 1344
Asp Arg Arg Ile Val Ser Lys Lys Tyr Lys Ile Tyr Phe Trp Asn Ile
435 440 445

atc acc att gct gtg ttt tac gcg ctg ccc gtg atc cag ctg gtc att 1392
Ile Thr Ile Ala Val Phe Tyr Ala Leu Pro Val Ile Gln Leu Val Ile
450 455 460

acc tat cag aca gtg gta aat gtc act ggc aac cag gac atc tgt tac 1440
Thr Tyr Gln Thr Val Val Asn Val Thr Gly Asn Gln Asp Ile Cys Tyr
465 470 475 480

tac aac ttc ctc tgt gct cac ccc ttg ggc gtc ctg agt gcc ttc aac 1488
Tyr Asn Phe Leu Cys Ala His Pro Leu Gly Val Leu Ser Ala Phe Asn
485 490 495

aac att ctc agc aat ctg ggc cac gtg ctt ctg ggc ttc ctc ttc ctg 1536
Asn Ile Leu Ser Asn Leu Gly His Val Leu Leu Gly Phe Leu Phe Leu
500 505 510

ctg ata gtc ttg cgc cgc gac atc ctc cat cgg aga gcc ctg gaa gcc 1584
Leu Ile Val Leu Arg Arg Asp Ile Leu His Arg Arg Ala Leu Glu Ala
515 520 525

aag gac atc ttt gct gtg gag tac ggg att ccc aaa cac ttt ggt ctc 1632
Lys Asp Ile Phe Ala Val Glu Tyr Gly Ile Pro Lys His Phe Gly Leu
530 535 540

ttc tac gct atg ggc att gca ttg atg atg gaa ggg gtg ctc agt gct 1680
Phe Tyr Ala Met Gly Ile Ala Leu Met Met Glu Gly Val Leu Ser Ala
545 550 555 560

tgc tac cat gtc tgc cct aat tat tcc aac ttc caa ttc gac acc tcc 1728
Cys Tyr His Val Cys Pro Asn Tyr Ser Asn Phe Gln Phe Asp Thr Ser
565 570 575

ttc atg tac atg atc gct ggc ctg tgc atg ctg aag ctc tat cag acc 1776
Phe Met Tyr Met Ile Ala Gly Leu Cys Met Leu Lys Leu Tyr Gln Thr
580 585 590

cgc cac cca gac atc aat gcc agc gcc tac tct gcc tat gcc tcc ttt 1824
Arg His Pro Asp Ile Asn Ala Ser Ala Tyr Ser Ala Tyr Ala Ser Phe
595 600 605

gct gtg gtc atc atg gtc acc gtc ctt gga gtg gtg ttt gga aaa aat 1872
Ala Val Val Ile Met Val Thr Val Leu Gly Val Val Phe Gly Lys Asn
610 615 620

gac gta tgg ttc tgg gtc atc ttc tct gca atc cac gtt ctg gcc tcg 1920
Asp Val Trp Phe Trp Val Ile Phe Ser Ala Ile His Val Leu Ala Ser
625 630 635 640

cta gcc ctc agc acc cag ata tat tat atg ggt cgt ttc aag ata gat 1968
Leu Ala Leu Ser Thr Gln Ile Tyr Tyr Met Gly Arg Phe Lys Ile Asp
645 650 655

ttg gga att ttc cgg cgg gct gcc atg gtg ttc tac aca gac tgt atc 2016
Leu Gly Ile Phe Arg Arg Ala Ala Met Val Phe Tyr Thr Asp Cys Ile
660 665 670

cag cag tgt agc cga cct cta tat atg gat aga atg gtg ttg ctg gtt 2064
Gln Gln Cys Ser Arg Pro Leu Tyr Met Asp Arg Met Val Leu Leu Val
675 680 685

gtg ggg aat ctg gtt aac tgg tcc ttc gcc ctc ttt gga ttg ata tac 2112
Val Gly Asn Leu Val Asn Trp Ser Phe Ala Leu Phe Gly Leu Ile Tyr
690 695 700

cgc ccc agg gac ttt gct tcc tac atg ctg ggc atc ttc atc tgt aac 2160
Arg Pro Arg Asp Phe Ala Ser Tyr Met Leu Gly Ile Phe Ile Cys Asn
705 710 715 720

ctt ttg ctg tac ctg gcc ttt tac atc atc atg aag ctc cgc agc tct 2208
Leu Leu Leu Tyr Leu Ala Phe Tyr Ile Ile Met Lys Leu Arg Ser Ser
725 730 735

gaa aag gtc ctc cca gtc ccg ctc ttc tgc atc gtg gcc acc gct gtg 2256
Glu Lys Val Leu Pro Val Pro Leu Phe Cys Ile Val Ala Thr Ala Val
740 745 750

atg tgg gct gcc gcc cta tat ttt ttc ttc cag aat ctc agc agc tgg 2304
Met Trp Ala Ala Ala Leu Tyr Phe Phe Phe Gln Asn Leu Ser Ser Trp
755 760 765

gag gga act ccg gcc gaa tcc cgg gag aag aac cgc gag tgc att ctg 2352
Glu Gly Thr Pro Ala Glu Ser Arg Glu Lys Asn Arg Glu Cys Ile Leu
770 775 780

ctg gat ttc ttc gat gac cat gac atc tgg cac ttc ctc tct gct act 2400
Leu Asp Phe Phe Asp Asp His Asp Ile Trp His Phe Leu Ser Ala Thr
785 790 795 800

gct ctg ttt ttc tca ttc ttg gtt ttg tta act ttg gat gat gac ctt 2448
Ala Leu Phe Phe Ser Phe Leu Val Leu Leu Thr Leu Asp Asp Asp Leu
805 810 815

gat gtg gtt cgg aga gac cag atc cct gtc ttc tga acctccaaca 2494
Asp Val Val Arg Arg Asp Gln Ile Pro Val Phe

820 825

ttaagagagg ggagggagcg atcaatcttg gtgctgtttc acaaaaatta cagtgaccac 2554

agcaaagtaa ccactgccag atgctccact caccctctgt agagccaggc tctgcattca 2614

cacaggaagg agaggggctg cgggagattt aaacctgcaa gaaaggaggc agaaggggag 2674

ccatgttttg aggacagacg caaacctgag gagctgagaa acacttgctc cttccatctg 2734

cagctttggg agtgcaacag ggataggcac tgcatccaag tcaactcacc atcttggggt 2794

ccctcccacc ctcacggaga cttgccagca atggcagaat gctgctgcac actttccttc 2854

aagtgtcacc ctgcccaaaa aaggccagca gcttggactt cctgcccaga aactgtgttg 2914

gccccccttca cacctctgca acacctgctg ctccagcaag aggatgtgat tctttagaat 2974

atggcgggga ggtgacccca ggccctgccc tactgggata gatgttttaa tggcaccagc 3034

tagtcacctc ccagaagaaa ctctgtatat ttcccccagg tttctgatgc catcagaagg 3094

gctcaggagt ggggtttgtc acacattcct cttaacaagt aactgtcact gggaccgagt 3154

cctgggtgct tacatattcc ttcgtgtctt catc 3188

<210> 4
<211> 827
<212> PRT
<213> Homo sapiens

<400> 4

Met Arg Gly Cys Leu Arg Leu Ala Leu Leu Cys Ala Leu Pro Trp Leu
1 5 10 15

Leu Leu Ala Ala Ser Pro Gly His Pro Ala Lys Ser Pro Arg Gln Pro
20 25 30

Pro Ala Pro Arg Arg Asp Pro Phe Asp Ala Ala Arg Gly Ala Asp Phe
35 40 45

Asp His Val Tyr Ser Gly Val Val Asn Leu Ser Thr Glu Asn Ile Tyr
50 55 60

Ser Phe Asn Tyr Thr Ser Gln Pro Asp Gln Val Thr Ala Val Arg Val
65 70 75 80

Tyr Val Asn Ser Ser Ser Glu Asn Leu Asn Tyr Pro Val Leu Val Val
85 90 95

Val Arg Gln Gln Lys Glu Val Leu Ser Trp Gln Val Pro Leu Leu Phe
100 105 110

Gln Gly Leu Tyr Gln Arg Ser Tyr Asn Tyr Gln Glu Val Ser Arg Thr

```
        115                 120                 125

Leu Cys Pro Ser Glu Ala Thr Asn Glu Thr Gly Pro Leu Gln Gln Leu
    130                 135                 140

Ile Phe Val Asp Val Ala Ser Met Ala Pro Leu Gly Ala Gln Tyr Lys
145                 150                 155                 160

Leu Leu Val Thr Lys Leu Lys His Phe Gln Leu Arg Thr Asn Val Ala
                165                 170                 175

Phe His Phe Thr Ala Ser Pro Ser Gln Pro Gln Tyr Phe Leu Tyr Lys
            180                 185                 190

Phe Pro Lys Asp Val Asp Ser Val Ile Ile Lys Val Val Ser Glu Met
        195                 200                 205

Ala Tyr Pro Cys Ser Val Val Ser Val Gln Asn Ile Met Cys Pro Val
    210                 215                 220

Tyr Asp Leu Asp His Asn Val Glu Phe Asn Gly Val Tyr Gln Ser Met
225                 230                 235                 240

Thr Lys Lys Ala Ala Ile Thr Leu Gln Lys Lys Asp Phe Pro Gly Glu
                245                 250                 255

Gln Phe Phe Val Val Phe Val Ile Lys Pro Glu Asp Tyr Ala Cys Gly
            260                 265                 270

Gly Ser Phe Phe Ile Gln Glu Lys Glu Asn Gln Thr Trp Asn Leu Gln
        275                 280                 285

Arg Lys Lys Asn Leu Glu Val Thr Ile Val Pro Ser Ile Lys Glu Ser
    290                 295                 300

Val Tyr Val Lys Ser Ser Leu Phe Ser Val Phe Ile Phe Leu Ser Phe
305                 310                 315                 320

Tyr Leu Gly Cys Leu Leu Val Gly Phe Val His Tyr Leu Arg Phe Gln
                325                 330                 335

Arg Lys Ser Ile Asp Gly Ser Phe Gly Ser Asn Asp Gly Ser Gly Asn
            340                 345                 350

Met Val Ala Ser His Pro Ile Ala Ala Ser Thr Pro Glu Gly Ser Asn
        355                 360                 365
```

Tyr Gly Thr Ile Asp Glu Ser Ser Ser Ser Pro Gly Arg Gln Met Ser
370 375 380

Ser Ser Asp Gly Gly Pro Pro Gly Gln Ser Asp Thr Asp Ser Ser Val
385 390 395 400

Glu Glu Ser Asp Phe Asp Thr Met Pro Asp Ile Glu Ser Asp Lys Asn
405 410 415

Ile Ile Arg Thr Lys Met Phe Leu Tyr Leu Ser Asp Leu Ser Arg Lys
420 425 430

Asp Arg Arg Ile Val Ser Lys Lys Tyr Lys Ile Tyr Phe Trp Asn Ile
435 440 445

Ile Thr Ile Ala Val Phe Tyr Ala Leu Pro Val Ile Gln Leu Val Ile
450 455 460

Thr Tyr Gln Thr Val Val Asn Val Thr Gly Asn Gln Asp Ile Cys Tyr
465 470 475 480

Tyr Asn Phe Leu Cys Ala His Pro Leu Gly Val Leu Ser Ala Phe Asn
485 490 495

Asn Ile Leu Ser Asn Leu Gly His Val Leu Leu Gly Phe Leu Phe Leu
500 505 510

Leu Ile Val Leu Arg Arg Asp Ile Leu His Arg Arg Ala Leu Glu Ala
515 520 525

Lys Asp Ile Phe Ala Val Glu Tyr Gly Ile Pro Lys His Phe Gly Leu
530 535 540

Phe Tyr Ala Met Gly Ile Ala Leu Met Met Glu Gly Val Leu Ser Ala
545 550 555 560

Cys Tyr His Val Cys Pro Asn Tyr Ser Asn Phe Gln Phe Asp Thr Ser
565 570 575

Phe Met Tyr Met Ile Ala Gly Leu Cys Met Leu Lys Leu Tyr Gln Thr
580 585 590

Arg His Pro Asp Ile Asn Ala Ser Ala Tyr Ser Ala Tyr Ala Ser Phe
595 600 605

Ala Val Val Ile Met Val Thr Val Leu Gly Val Val Phe Gly Lys Asn
610 615 620

Asp Val Trp Phe Trp Val Ile Phe Ser Ala Ile His Val Leu Ala Ser
625 630 635 640

Leu Ala Leu Ser Thr Gln Ile Tyr Tyr Met Gly Arg Phe Lys Ile Asp
645 650 655

Leu Gly Ile Phe Arg Arg Ala Ala Met Val Phe Tyr Thr Asp Cys Ile
660 665 670

Gln Gln Cys Ser Arg Pro Leu Tyr Met Asp Arg Met Val Leu Leu Val
675 680 685

Val Gly Asn Leu Val Asn Trp Ser Phe Ala Leu Phe Gly Leu Ile Tyr
690 695 700

Arg Pro Arg Asp Phe Ala Ser Tyr Met Leu Gly Ile Phe Ile Cys Asn
705 710 715 720

Leu Leu Leu Tyr Leu Ala Phe Tyr Ile Ile Met Lys Leu Arg Ser Ser
725 730 735

Glu Lys Val Leu Pro Val Pro Leu Phe Cys Ile Val Ala Thr Ala Val
740 745 750

Met Trp Ala Ala Ala Leu Tyr Phe Phe Phe Gln Asn Leu Ser Ser Trp
755 760 765

Glu Gly Thr Pro Ala Glu Ser Arg Glu Lys Asn Arg Glu Cys Ile Leu
770 775 780

Leu Asp Phe Phe Asp Asp His Asp Ile Trp His Phe Leu Ser Ala Thr
785 790 795 800

Ala Leu Phe Phe Ser Phe Leu Val Leu Leu Thr Leu Asp Asp Asp Leu
805 810 815

Asp Val Val Arg Arg Asp Gln Ile Pro Val Phe
820 825

<210> 5
<211> 2484
<212> DNA
<213> Homo sapiens

<400> 5
atgcgcggct gcctgcggct cgcgctgctc tgcgcgctgc cctggctcct gctggcggcg 60

tcgcccgggc acccggcgaa atccccccagg cagcccccgg caccgcgccg cgaccccttc 120

gacgctgcca ggggcgccga tttcgatcat gtctacagcg gggtggtgaa cctcagcacc 180
gagaacatct actctttcaa ctacaccagc cagcccgacc aggtgacagc cgtgagggtg 240
tatgtgaaca gttcctctga gaatctcaac tacccggtcc ttgttgtggt tcgccagcag 300
aaagaggtgc tgtcctggca ggttcctctg ctcttccaag gactatacca gaggagctac 360
aattatcaag aagtgagccg caccttatgt ccctcagaag caaccaatga gacgggaccc 420
ttgcagcaac tgatatttgt agatgtcgca tccatggcac ccctgggtgc tcagtacaaa 480
ctgctagtta ccaagctgaa gcacttccag ctccggacaa atgttgcctt tcactttact 540
gccagcccct ctcaacctca gtattttcta tacaagtttc ccaaagacgt ggactcagtt 600
atcattaaag tggtgtctga aatggcttat ccatgttctg ttgtctcagt ccagaatatc 660
atgtgcccgg tgtatgatct cgaccacaat gtggaattta atggtgtcta tcagtccatg 720
accaagaaag ctgccatcac gctacagaag aaggattttc caggcgagca gttcttcgtg 780
gtatttgtga taaagcctga agattatgcc tgtggaggat ctttcttcat ccaggaaaag 840
gaaaaccaga cctggaatct acagcgaaaa aagaaccttg aagtgaccat tgtcccttcc 900
attaaagaat ctgtttatgt gaaatccagt cttttcagtg tcttcatctt cctgtccttc 960
tacttgggat gccttcttgt tgggtttgtt cattatctga ggtttcagag aaaatccatt 1020
gatggaagct ttgggtccaa tgatggctct ggaaatatgg tggcatctca tcccattgct 1080
gccagcacac ccgaagggag caattatggg acaatagatg agtcaagctc cagtcctgga 1140
aggcagatgt cctcctccga tggtgggcca ccgggccagt cagacacaga cagctccgtg 1200
gaggagagcg acttcgacac catgccagac attgagagtg ataaaaacat catccggacc 1260
aagatgttcc tttacctgtc agatttgtcc aggaaggacc ggagaattgt cagcaaaaaa 1320
tataaaattt atttttggaa catcatcacc attgctgtgt tttacgcgct gcccgtgatc 1380
cagctggtca ttacctatca gacagtggta aatgtcactg gcaaccagga catctgttac 1440
tacaacttcc tctgtgctca cccccttgggc gtcctgagtg ccttcaacaa cattctcagc 1500
aatctgggcc acgtgcttct gggcttcctc ttcctgctga tagtcttgcg ccgcgacatc 1560
ctccatcgga gagccctgga agccaaggac atctttgctg tggagtacgg gattcccaaa 1620
cactttggtc tcttctacgc tatgggcatt gcattgatga tggaaggggt gctcagtgct 1680
tgctaccatg tctgccctaa ttattccaac ttccaattcg acacctcctt catgtacatg 1740
atcgctggcc tgtgcatgct gaagctctat cagacccgcc acccagacat caatgccagc 1800
gcctactctg cctatgcctc ctttgctgtg gtcatcatgg tcaccgtcct tggagtggtg 1860
tttggaaaaa atgacgtatg gttctgggtc atcttctctg caatccacgt tctggcctcg 1920
ctagccctca gcacccagat atattatatg ggtcgtttca agatagattt gggaattttc 1980
cggcgggctg ccatggtgtt ctacacagac tgtatccagc agtgtagccg acctctatat 2040

atggatagaa tggtgttgct ggttgtgggg aatctggtta actggtcctt cgccctcttt 2100

ggattgatat accgccccag ggactttgct tcctacatgc tgggcatctt catctgtaac 2160

cttttgctgt acctggcctt ttacatcatc atgaagctcc gcagctctga aaaggtcctc 2220

ccagtcccgc tcttctgcat cgtggccacc gctgtgatgt gggctgccgc cctatatttt 2280

ttcttccaga atctcagcag ctgggaggga actccggccg aatcccggga gaagaaccgc 2340

gagtgcattc tgctggattt cttcgatgac catgacatct ggcacttcct ctctgctact 2400

gctctgtttt tctcattctt ggttttgtta actttggatg atgaccttga tgtggttcgg 2460

agagaccaga tccctgtctt ctga 2484

<210> 6
<211> 108
<212> PRT
<213> Homo sapiens

<400> 6

Val Ile Phe Ser Ala Ile His Val Leu Ala Ser Leu Ala Leu Ser Thr
1 5 10 15

Gln Ile Tyr Tyr Met Gly Arg Phe Lys Ile Asp Leu Gly Ile Phe Arg
20 25 30

Arg Ala Ala Met Val Phe Tyr Thr Asp Cys Ile Gln Gln Cys Ser Arg
35 40 45

Pro Leu Tyr Met Asp Arg Met Val Leu Leu Val Val Gly Asn Leu Val
50 55 60

Asn Trp Ser Phe Ala Leu Phe Gly Leu Ile Tyr Arg Pro Arg Asp Phe
65 70 75 80

Ala Ser Tyr Met Leu Gly Ile Phe Ile Cys Asn Leu Leu Leu Tyr Leu
85 90 95

Ala Phe Tyr Ile Ile Met Lys Leu Arg Ser Ser Glu
100 105

<210> 7
<211> 19
<212> PRT
<213> Homo sapiens

<400> 7

Met Arg Gly Cys Leu Arg Leu Ala Leu Leu Cys Ala Leu Pro Trp Leu
1 5 10 15

Leu Leu Ala

<210> 8
<211> 808
<212> PRT
<213> Homo sapiens

<400> 8

Ala Ser Pro Gly His Pro Ala Lys Ser Pro Arg Gln Pro Pro Ala Pro
1 5 10 15

Arg Arg Asp Pro Phe Asp Ala Ala Arg Gly Ala Asp Phe Asp His Val
20 25 30

Tyr Ser Gly Val Val Asn Leu Ser Thr Glu Asn Ile Tyr Ser Phe Asn
35 40 45

Tyr Thr Ser Gln Pro Asp Gln Val Thr Ala Val Arg Val Tyr Val Asn
50 55 60

Ser Ser Ser Glu Asn Leu Asn Tyr Pro Val Leu Val Val Val Arg Gln
65 70 75 80

Gln Lys Glu Val Leu Ser Trp Gln Val Pro Leu Leu Phe Gln Gly Leu
85 90 95

Tyr Gln Arg Ser Tyr Asn Tyr Gln Glu Val Ser Arg Thr Leu Cys Pro
100 105 110

Ser Glu Ala Thr Asn Glu Thr Gly Pro Leu Gln Gln Leu Ile Phe Val
115 120 125

Asp Val Ala Ser Met Ala Pro Leu Gly Ala Gln Tyr Lys Leu Leu Val
130 135 140

Thr Lys Leu Lys His Phe Gln Leu Arg Thr Asn Val Ala Phe His Phe
145 150 155 160

Thr Ala Ser Pro Ser Gln Pro Gln Tyr Phe Leu Tyr Lys Phe Pro Lys
165 170 175

Asp Val Asp Ser Val Ile Ile Lys Val Val Ser Glu Met Ala Tyr Pro
180 185 190

Cys Ser Val Val Ser Val Gln Asn Ile Met Cys Pro Val Tyr Asp Leu
195 200 205

Asp His Asn Val Glu Phe Asn Gly Val Tyr Gln Ser Met Thr Lys Lys
210 215 220

Ala Ala Ile Thr Leu Gln Lys Lys Asp Phe Pro Gly Glu Gln Phe Phe
225 230 235 240

Val Val Phe Val Ile Lys Pro Glu Asp Tyr Ala Cys Gly Gly Ser Phe
245 250 255

Phe Ile Gln Glu Lys Glu Asn Gln Thr Trp Asn Leu Gln Arg Lys Lys
260 265 270

Asn Leu Glu Val Thr Ile Val Pro Ser Ile Lys Glu Ser Val Tyr Val
275 280 285

Lys Ser Ser Leu Phe Ser Val Phe Ile Phe Leu Ser Phe Tyr Leu Gly
290 295 300

Cys Leu Leu Val Gly Phe Val His Tyr Leu Arg Phe Gln Arg Lys Ser
305 310 315 320

Ile Asp Gly Ser Phe Gly Ser Asn Asp Gly Ser Gly Asn Met Val Ala
325 330 335

Ser His Pro Ile Ala Ala Ser Thr Pro Glu Gly Ser Asn Tyr Gly Thr
340 345 350

Ile Asp Glu Ser Ser Ser Ser Pro Gly Arg Gln Met Ser Ser Ser Asp
355 360 365

Gly Gly Pro Pro Gly Gln Ser Asp Thr Asp Ser Ser Val Glu Glu Ser
370 375 380

Asp Phe Asp Thr Met Pro Asp Ile Glu Ser Asp Lys Asn Ile Ile Arg
385 390 395 400

Thr Lys Met Phe Leu Tyr Leu Ser Asp Leu Ser Arg Lys Asp Arg Arg
405 410 415

Ile Val Ser Lys Lys Tyr Lys Ile Tyr Phe Trp Asn Ile Ile Thr Ile
420 425 430

Ala Val Phe Tyr Ala Leu Pro Val Ile Gln Leu Val Ile Thr Tyr Gln
435 440 445

Thr Val Val Asn Val Thr Gly Asn Gln Asp Ile Cys Tyr Tyr Asn Phe
450 455 460

Leu Cys Ala His Pro Leu Gly Val Leu Ser Ala Phe Asn Asn Ile Leu
465 470 475 480

Ser Asn Leu Gly His Val Leu Leu Gly Phe Leu Phe Leu Leu Ile Val
485 490 495

Leu Arg Arg Asp Ile Leu His Arg Arg Ala Leu Glu Ala Lys Asp Ile
500 505 510

Phe Ala Val Glu Tyr Gly Ile Pro Lys His Phe Gly Leu Phe Tyr Ala
515 520 525

Met Gly Ile Ala Leu Met Met Glu Gly Val Leu Ser Ala Cys Tyr His
530 535 540

Val Cys Pro Asn Tyr Ser Asn Phe Gln Phe Asp Thr Ser Phe Met Tyr
545 550 555 560

Met Ile Ala Gly Leu Cys Met Leu Lys Leu Tyr Gln Thr Arg His Pro
565 570 575

Asp Ile Asn Ala Ser Ala Tyr Ser Ala Tyr Ala Ser Phe Ala Val Val
580 585 590

Ile Met Val Thr Val Leu Gly Val Val Phe Gly Lys Asn Asp Val Trp
595 600 605

Phe Trp Val Ile Phe Ser Ala Ile His Val Leu Ala Ser Leu Ala Leu
610 615 620

Ser Thr Gln Ile Tyr Tyr Met Gly Arg Phe Lys Ile Asp Leu Gly Ile
625 630 635 640

Phe Arg Arg Ala Ala Met Val Phe Tyr Thr Asp Cys Ile Gln Gln Cys
645 650 655

Ser Arg Pro Leu Tyr Met Asp Arg Met Val Leu Leu Val Val Gly Asn
660 665 670

Leu Val Asn Trp Ser Phe Ala Leu Phe Gly Leu Ile Tyr Arg Pro Arg
675 680 685

Asp Phe Ala Ser Tyr Met Leu Gly Ile Phe Ile Cys Asn Leu Leu Leu
690 695 700

Tyr Leu Ala Phe Tyr Ile Ile Met Lys Leu Arg Ser Ser Glu Lys Val
705 710 715 720

Leu Pro Val Pro Leu Phe Cys Ile Val Ala Thr Ala Val Met Trp Ala
725 730 735

Ala Ala Leu Tyr Phe Phe Phe Gln Asn Leu Ser Ser Trp Glu Gly Thr
740 745 750

Pro Ala Glu Ser Arg Glu Lys Asn Arg Glu Cys Ile Leu Leu Asp Phe
755 760 765

Phe Asp Asp His Asp Ile Trp His Phe Leu Ser Ala Thr Ala Leu Phe
770 775 780

Phe Ser Phe Leu Val Leu Leu Thr Leu Asp Asp Asp Leu Asp Val Val
785 790 795 800

Arg Arg Asp Gln Ile Pro Val Phe
805

<210> 9
<211> 128
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(128)
<223> X = any amino acid or a stop codon

<400> 9

Val Gly Glu Pro Tyr Ile Asp Trp Asp Glu Phe Pro Glu Leu Leu Ser
1 5 10 15

Arg Thr Ala Val Arg Ala Arg Lys Ile Pro Ile Ser Asp Thr Ile Xaa
20 25 30

Lys Thr Lys Ala Lys Gln Val Val Lys Leu Leu Ser Asn Ile Arg Ser
35 40 45

Gln Ala Val Gly Ile Leu Met Ser Ser Leu His Leu Asp Met Lys Asp
50 55 60

Ile Gln His Ala Val Val Asn Leu Asp Asn Ser Val Val Asp Leu Glu
65 70 75 80

Thr Leu Gln Ala Leu Tyr Glu Asn Arg Ala Gln Ser Asp Glu Leu Glu
85 90 95

Xaa Ile Glu Lys His Gly Arg Ser Ser Lys Asp Lys Glu Asn Ala Lys
100 105 110

Ser Leu Asp Lys Pro Glu Gln Leu Tyr Phe Leu Arg Phe Leu Tyr Glu
115 120 125

<210> 10
<211> 391
<212> DNA
<213> Homo sapiens

<400> 10
tgcagcccta tgtgcaggaa agccagaaca agttcttcaa gctgcccgtg tccgtggtca 60

acaccacact gccgtgcgtg gcctacgtgc tgctgtcact cgtgtacttg cccggcgtgc 120

tggcggctgc cctgcagctg cggcgcggca ccaagtacca gcgcttcccc gactggctgg 180

accactggct acagcaccgc aagcagatcg ggctgctcag cttcttctgc gccgccctgc 240

acgccctcta cagcttctgc ttgccgctgc gccgcgccca ccgctacgac ctggtcaacc 300

tggcagtcaa gcaggtcttg gccaacaaga gccacctctg ggtggaggaa gaggtctggc 360

ggatggagat ctacctctcc ctgggagtgc t 391

<210> 11
<211> 1169
<212> DNA
<213> Homo sapiens

<400> 11
tgcagcccta tgtgcaggaa agccagaaca agttcttcaa gctgcccgtg tccgtggtca 60

acaccacact gccgtgcgtg gcctacgtgc tgctgtcact cgtgtacttg cccggcgtgc 120

tggcggctgc cctgcagctg cggcgcggca ccaagtacca gcgcttcccc gactggctgg 180

accactggct acagcaccgc aagcagatcg ggctgctcag cttcttctgc gccgccctgc 240

acgccctcta cagcttctgc ttgccgctgc gccgcgccca ccgctacgac ctggtcaacc 300

tggcagtcaa gcaggtcttg gccaacaaga gccacctctg ggtggaggag gaggtctggc 360

ggatggagat ctacctctcc ctgggagtgc tggccctcgg cacgttgtcc ctgctggccg 420

tgacctcact gccgtccatt gcaaactcgc tcaactggag ggagttcagc ttcgttcagt 480

cctcactggg ctttgtggcc ctcgtgctga gcacactgca cacgctcacc tacggctgga 540

cccgcgcctt cgaggagagc cgctacaagt tctacctgcc tcccaccttc acgctcacgc 600

tgctggtgcc ctgcgtcgtc atcctggcca aagccctgtt tctcctgccc tgcatcagcc 660

gcagactcgc caggatccgg agaggctggg agagggagag caccatcaag ttcacgctgc 720

ccacagacca cgccctggcc gagaagacga gccacgtatg aggtgcctgc cctgggctct 780

ggaccccggg cacacgaggg acggtgccct gagcccgtta ggttttcttt tcttggtggt 840

gcaaagtggt ataactgtgt gcaaatagga ggtttgaggt ccaaattcct gggactcaaa 900

tgtatgcagt actattcaga atgatataca cacatatgtg tatatgtatt tacatatatt 960

ccacatatat aacaggattt gcaattatac atagctagct aaaaagttgg gtctctgaga 1020

tttcaacttg tagatttaaa aacaagtgcc gtacgttaag agaagagcag atcatgctat 1080

tgtgacattt gcagagatat acacacactt ttttgtacag aagaggcttg tgctgtggtg 1140

ggttcgattt atccctgccc accccatcc 1169

<210> 12
<211> 2936
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (135)..(1601)

<400> 12
accgccttcg ccgcggacct tcagctgccg cggtcgctcc gagcggcggg ccgcagaggt 60

tcaagcgatt ctcctgcttc agcctccgga gtagctggga ttacaggcac gtgccaacac 120

```
acccagccac caaa atg cca gaa gag atg gac aag cca ctg atc agc ctc      170
                Met Pro Glu Glu Met Asp Lys Pro Leu Ile Ser Leu
                1               5                   10

cac ctg gtg gac agc gat agt agc ctt gcc aag gtc ccc gat gag gcc      218
His Leu Val Asp Ser Asp Ser Ser Leu Ala Lys Val Pro Asp Glu Ala
        15                  20                  25

ccc aaa gtg ggc atc ctg ggt agc ggg gac ttt gcc cgc tcc ctg gcc      266
Pro Lys Val Gly Ile Leu Gly Ser Gly Asp Phe Ala Arg Ser Leu Ala
    30                  35                  40

aca cgc ctg gtg ggc tct ggc ttc aaa gtg gtg gtg ggg agc cgc aac      314
Thr Arg Leu Val Gly Ser Gly Phe Lys Val Val Val Gly Ser Arg Asn
45                  50                  55                  60

ccc aaa cgc aca gcc agg ctg tat ccc tca gcg gcc caa gtg act ttc      362
Pro Lys Arg Thr Ala Arg Leu Tyr Pro Ser Ala Ala Gln Val Thr Phe
                65                  70                  75

caa gag gag gca gtg agc tcc ccg gag gtc atc ttt gtg gct gtg ttc      410
Gln Glu Glu Ala Val Ser Ser Pro Glu Val Ile Phe Val Ala Val Phe
            80                  85                  90

cgg gag cac tac tct tca ctg tgc agt ctc agt gac cag ctg gcg ggc      458
Arg Glu His Tyr Ser Ser Leu Cys Ser Leu Ser Asp Gln Leu Ala Gly
        95                  100                 105

aag atc ctg gtg gat gtg agc aac cct aca gag caa gag cac ctt cag      506
Lys Ile Leu Val Asp Val Ser Asn Pro Thr Glu Gln Glu His Leu Gln
    110                 115                 120

cat cgt gag tcc aat gct gag tac ctg gcc tcc ctc ttc ccc act tgc      554
His Arg Glu Ser Asn Ala Glu Tyr Leu Ala Ser Leu Phe Pro Thr Cys
125                 130                 135                 140

aca gtg gtc aag gcc ttc aat gtc atc tct gcc tgg acc ctg cag gct      602
Thr Val Val Lys Ala Phe Asn Val Ile Ser Ala Trp Thr Leu Gln Ala
                145                 150                 155
```

ggc cca agg gat ggt aac agg cag gtg ccc atc tgc ggt gac cag cca 650
Gly Pro Arg Asp Gly Asn Arg Gln Val Pro Ile Cys Gly Asp Gln Pro
160 165 170

gaa gcc aag cgt gct gtc tcg gag atg gcg ctc gcc atg ggc ttc atg 698
Glu Ala Lys Arg Ala Val Ser Glu Met Ala Leu Ala Met Gly Phe Met
175 180 185

ccc gtg gac atg gga tcc ctg gcg tca gcc tgg gag gtg gag gcc atg 746
Pro Val Asp Met Gly Ser Leu Ala Ser Ala Trp Glu Val Glu Ala Met
190 195 200

ccc ctg cgc ctc ctc ccg gcc tgg aag gtg ccc acc ctg ctg gcc ctg 794
Pro Leu Arg Leu Leu Pro Ala Trp Lys Val Pro Thr Leu Leu Ala Leu
205 210 215 220

ggg ctc ttc gtc tgc ttc tat gcc tac aac ttc gtc cgg gac gtt ctg 842
Gly Leu Phe Val Cys Phe Tyr Ala Tyr Asn Phe Val Arg Asp Val Leu
225 230 235

cag ccc tat gtg cag gaa agc cag aac aag ttc ttc aag ctg ccc gtg 890
Gln Pro Tyr Val Gln Glu Ser Gln Asn Lys Phe Phe Lys Leu Pro Val
240 245 250

tcc gtg gtc aac acc aca ctg ccg tgc gtg gcc tac gtg ctg ctg tca 938
Ser Val Val Asn Thr Thr Leu Pro Cys Val Ala Tyr Val Leu Leu Ser
255 260 265

ctc gtg tac ttg ccc ggc gtg ctg gcg gct gcc ctg cag ctg cgg cgc 986
Leu Val Tyr Leu Pro Gly Val Leu Ala Ala Ala Leu Gln Leu Arg Arg
270 275 280

ggc acc aag tac cag cgc ttc ccc gac tgg ctg gac cac tgg cta cag 1034
Gly Thr Lys Tyr Gln Arg Phe Pro Asp Trp Leu Asp His Trp Leu Gln
285 290 295 300

cac cgc aag cag atc ggg ctg ctc agc ttc ttc tgc gcc gcc ctg cac 1082
His Arg Lys Gln Ile Gly Leu Leu Ser Phe Phe Cys Ala Ala Leu His
305 310 315

gcc ctc tac agc ttc tgc ttg ccg ctg cgc cgc gcc cac cgc tac gac 1130
Ala Leu Tyr Ser Phe Cys Leu Pro Leu Arg Arg Ala His Arg Tyr Asp
320 325 330

ctg gtc aac ctg gca gtc aag cag gtc ttg gcc aac aag agc cac ctc 1178
Leu Val Asn Leu Ala Val Lys Gln Val Leu Ala Asn Lys Ser His Leu
335 340 345

tgg gtg gag gag gag gtc tgg cgg atg gag atc tac ctc tcc ctg gga 1226
Trp Val Glu Glu Glu Val Trp Arg Met Glu Ile Tyr Leu Ser Leu Gly
350 355 360

gtg ctg gcc ctc ggc acg ttg tcc ctg ctg gcc gtg acc tca ctg ccg 1274
Val Leu Ala Leu Gly Thr Leu Ser Leu Leu Ala Val Thr Ser Leu Pro
365 370 375 380

tcc att gca aac tcg ctc aac tgg agg gag ttc agc ttc gtt cag tcc 1322
Ser Ile Ala Asn Ser Leu Asn Trp Arg Glu Phe Ser Phe Val Gln Ser
385 390 395

tca ctg ggc ttt gtg gcc ctc gtg ctg agc aca ctg cac acg ctc acc 1370
Ser Leu Gly Phe Val Ala Leu Val Leu Ser Thr Leu His Thr Leu Thr
400 405 410

```
tac ggc tgg acc cgc gcc ttc gag gag agc cgc tac aag ttc tac ctg      1418
Tyr Gly Trp Thr Arg Ala Phe Glu Glu S r Arg Tyr Lys Phe Tyr Leu
        415                 420                 425

cct ccc acc ttc acg ctc acg ctg ctg gtg ccc tgc gtc gtc atc ctg      1466
Pro Pro Thr Phe Thr Leu Thr Leu Leu Val Pro Cys Val Val Ile Leu
    430                 435                 440

gcc aaa gcc ctg ttt ctc ctg ccc tgc atc agc cgc aga ctc gcc agg      1514
Ala Lys Ala Leu Phe Leu Leu Pro Cys Ile Ser Arg Arg Leu Ala Arg
445                 450                 455                 460

atc cgg aga ggc tgg gag agg gag agc acc atc aag ttc acg ctg ccc      1562
Ile Arg Arg Gly Trp Glu Arg Glu Ser Thr Ile Lys Phe Thr Leu Pro
            465                 470                 475

aca gac cac gcc ctg gcc gag aag acg agc cac gta tga ggtgcctgcc       1611
Thr Asp His Ala Leu Ala Glu Lys Thr Ser His Val
        480                 485

ctgggctctg gaccccgggc acacgaggga cggtgccctg agcccgttag gttttctttt   1671

cttggtggtg caaagtggta taactgtgtg caaataggag gtttgaggtc caaattcctg   1731

ggactcaaat gtatgcagta ctattcagaa tgatatacac acatatgtgt atatgtattt   1791

acatatattc cacatatata acaggatttg caattataca tagctagcta aaaagttggg   1851

tctctgagat ttcaacttgt agatttaaaa acaagtgccg tacgttaaga gaagagcaga   1911

tcatgctatt gtgacatttg cagagatata cacacacttt ttgtacagaa gaggcttgtg   1971

ctgtggtggg ttcgatttat ccctgcccac cccatcccca caacttccct tttgctactt   2031

ccccaaggct cttgcagagc tagggctctg aaggggaggg aaggcaacgg ctctgcccag   2091

agccatccct ggagcatgtg agcagcggct ggtctcttcc ctccacctgg ggcagcagca   2151

ggaggcctgg gggggaggaa aatcaggcag tcggcctgga gtctgtgcct ggtcctttgc   2211

ccggtggtgg gaggatggag ggattgggct gaagctgctc cacctcatcc ttgctgagtg   2271

ggggagacat tttccctgaa agtcagaagt caccatagag cctgcaaatg gatcctcctg   2331

tgagagtgac gtcacctcct ttccagagcc attagtgagc ctggcttggg aacaagtgta   2391

atttccttcc ctcctttaac ctggcgatga gcgtccttta aaccactgtg ccttctcacc   2451

ctttccatct tcagtttgaa cgactcccag gaaggcctag agcagaccct ttagaaatca   2511

gcccaagggg gagagcaaga gaaaacactc tagggagtaa agctccccgg gcgtcagagt   2571

tgagccctgc ctgggctgaa ggactgtctt cacgaagtca gtcctgagga aaaatattgg   2631

ggactccaaa tgtcctctgg cagaggaccc agaaaaccac actggctcca acttcctcct   2691

catggggcat tacacttcaa aacagtgggg agcaactttt ccaccaaagc tacaaaccta   2751

aaatgctgct gccccaaagc acaagaggga agagcaccgc cggggccaca ggacgtctgt   2811

cctccagtca caggccatcc ttgctgctcc ctactgactc tagcttactt cccctgtgaa   2871
```

gaaacaggtg ttctcggctg agcccccaac cctctgcaga accaggttga tctgccacag 2931

aaaaa 2936

<210> 13
<211> 488
<212> PRT
<213> Homo sapiens

<400> 13

Met Pro Glu Glu Met Asp Lys Pro Leu Ile Ser Leu His Leu Val Asp
1 5 10 15

Ser Asp Ser Ser Leu Ala Lys Val Pro Asp Glu Ala Pro Lys Val Gly
20 25 30

Ile Leu Gly Ser Gly Asp Phe Ala Arg Ser Leu Ala Thr Arg Leu Val
35 40 45

Gly Ser Gly Phe Lys Val Val Val Gly Ser Arg Asn Pro Lys Arg Thr
50 55 60

Ala Arg Leu Tyr Pro Ser Ala Ala Gln Val Thr Phe Gln Glu Glu Ala
65 70 75 80

Val Ser Ser Pro Glu Val Ile Phe Val Ala Val Phe Arg Glu His Tyr
85 90 95

Ser Ser Leu Cys Ser Leu Ser Asp Gln Leu Ala Gly Lys Ile Leu Val
100 105 110

Asp Val Ser Asn Pro Thr Glu Gln Glu His Leu Gln His Arg Glu Ser
115 120 125

Asn Ala Glu Tyr Leu Ala Ser Leu Phe Pro Thr Cys Thr Val Val Lys
130 135 140

Ala Phe Asn Val Ile Ser Ala Trp Thr Leu Gln Ala Gly Pro Arg Asp
145 150 155 160

Gly Asn Arg Gln Val Pro Ile Cys Gly Asp Gln Pro Glu Ala Lys Arg
165 170 175

Ala Val Ser Glu Met Ala Leu Ala Met Gly Phe Met Pro Val Asp Met
180 185 190

Gly Ser Leu Ala Ser Ala Trp Glu Val Glu Ala Met Pro Leu Arg Leu
195 200 205

Leu Pro Ala Trp Lys Val Pro Thr Leu Leu Ala Leu Gly Leu Phe Val
210 215 220

Cys Phe Tyr Ala Tyr Asn Phe Val Arg Asp Val Leu Gln Pro Tyr Val
225 230 235 240

Gln Glu Ser Gln Asn Lys Phe Phe Lys Leu Pro Val Ser Val Val Asn
245 250 255

Thr Thr Leu Pro Cys Val Ala Tyr Val Leu Leu Ser Leu Val Tyr Leu
260 265 270

Pro Gly Val Leu Ala Ala Ala Leu Gln Leu Arg Arg Gly Thr Lys Tyr
275 280 285

Gln Arg Phe Pro Asp Trp Leu Asp His Trp Leu Gln His Arg Lys Gln
290 295 300

Ile Gly Leu Leu Ser Phe Phe Cys Ala Ala Leu His Ala Leu Tyr Ser
305 310 315 320

Phe Cys Leu Pro Leu Arg Arg Ala His Arg Tyr Asp Leu Val Asn Leu
325 330 335

Ala Val Lys Gln Val Leu Ala Asn Lys Ser His Leu Trp Val Glu Glu
340 345 350

Glu Val Trp Arg Met Glu Ile Tyr Leu Ser Leu Gly Val Leu Ala Leu
355 360 365

Gly Thr Leu Ser Leu Leu Ala Val Thr Ser Leu Pro Ser Ile Ala Asn
370 375 380

Ser Leu Asn Trp Arg Glu Phe Ser Phe Val Gln Ser Ser Leu Gly Phe
385 390 395 400

Val Ala Leu Val Leu Ser Thr Leu His Thr Leu Thr Tyr Gly Trp Thr
405 410 415

Arg Ala Phe Glu Glu Ser Arg Tyr Lys Phe Tyr Leu Pro Pro Thr Phe
420 425 430

Thr Leu Thr Leu Leu Val Pro Cys Val Val Ile Leu Ala Lys Ala Leu
435 440 445

Phe Leu Leu Pro Cys Ile Ser Arg Arg Leu Ala Arg Ile Arg Arg Gly
450 455 460

Trp Glu Arg Glu Ser Thr Ile Lys Phe Thr Leu Pro Thr Asp His Ala
465 470 475 480

Leu Ala Glu Lys Thr Ser His Val
485

<210> 14
<211> 1467
<212> DNA
<213> Homo sapiens

<400> 14
atgccagaag agatggacaa gccactgatc agcctccacc tggtggacag cgatagtagc 60

cttgccaagg tccccgatga ggcccccaaa gtgggcatcc tgggtagcgg ggactttgcc 120

cgctccctgg ccacacgcct ggtgggctct ggcttcaaag tggtggtggg gagccgcaac 180

cccaaacgca cagccaggct gtatccctca gcggcccaag tgactttcca agaggaggca 240

gtgagctccc cggaggtcat ctttgtggct gtgttccggg agcactactc ttcactgtgc 300

agtctcagtg accagctggc gggcaagatc ctggtggatg tgagcaaccc tacagagcaa 360

gagcaccttc agcatcgtga gtccaatgct gagtacctgg cctccctctt ccccacttgc 420

acagtggtca aggccttcaa tgtcatctct gcctggaccc tgcaggctgg cccaagggat 480

ggtaacaggc aggtgcccat ctgcggtgac cagccagaag ccaagcgtgc tgtctcggag 540

atggcgctcg ccatgggctt catgcccgtg gacatgggat ccctggcgtc agcctgggag 600

gtggaggcca tgcccctgcg cctcctcccg gcctggaagg tgcccaccct gctggccctg 660

gggctcttcg tctgcttcta tgcctacaac ttcgtccggg acgttctgca gccctatgtg 720

caggaaagcc agaacaagtt cttcaagctg cccgtgtccg tggtcaacac cacactgccg 780

tgcgtggcct acgtgctgct gtcactcgtg tacttgcccg gcgtgctggc ggctgccctg 840

cagctgcggc gcggcaccaa gtaccagcgc ttccccgact ggctggacca ctggctacag 900

caccgcaagc agatcgggct gctcagcttc ttctgcgccg ccctgcacgc cctctacagc 960

ttctgcttgc cgctgcgccg cgcccaccgc tacgacctgg tcaacctggc agtcaagcag 1020

gtcttggcca acaagagcca cctctgggtg gaggaggagg tctggcggat ggagatctac 1080

ctctccctgg gagtgctggc cctcggcacg ttgtccctgc tggccgtgac ctcactgccg 1140

tccattgcaa actcgctcaa ctggagggag ttcagcttcg ttcagtcctc actgggcttt 1200

gtggccctcg tgctgagcac actgcacacg ctcacctacg gctggacccg cgccttcgag 1260

gagagccgct acaagttcta cctgcctccc accttcacgc tcacgctgct ggtgccctgc 1320

gtcgtcatcc tggccaaagc cctgtttctc ctgccctgca tcagccgcag actcgccagg 1380

atccggagag gctgggagag ggagagcacc atcaagttca cgctgcccac agaccacgcc 1440

ctggccgaga agacgagcca cgtatga 1467

<210> 15
<211> 237
<212> PRT
<213> Homo sapiens

<400> 15

Val Pro Thr Leu Leu Ala Leu Gly Leu Phe Val Cys Phe Tyr Ala Tyr
1 5 10 15

Asn Phe Val Arg Asp Val Leu Gln Pro Tyr Val Gln Glu Ser Gln Asn
20 25 30

Lys Phe Phe Lys Leu Pro Val Ser Val Val Asn Thr Thr Leu Pro Cys
35 40 45

Val Ala Tyr Val Leu Leu Ser Leu Val Tyr Leu Pro Gly Val Leu Ala
50 55 60

Ala Ala Leu Gln Leu Arg Arg Gly Thr Lys Tyr Gln Arg Phe Pro Asp
65 70 75 80

Trp Leu Asp His Trp Leu Gln His Arg Lys Gln Ile Gly Leu Leu Ser
85 90 95

Phe Phe Cys Ala Ala Leu His Ala Leu Tyr Ser Phe Cys Leu Pro Leu
100 105 110

Arg Arg Ala His Arg Tyr Asp Leu Val Asn Leu Ala Val Lys Gln Val
115 120 125

Leu Ala Asn Lys Ser His Leu Trp Val Glu Glu Glu Val Trp Arg Met
130 135 140

Glu Ile Tyr Leu Ser Leu Gly Val Leu Ala Leu Gly Thr Leu Ser Leu
145 150 155 160

Leu Ala Val Thr Ser Leu Pro Ser Ile Ala Asn Ser Leu Asn Trp Arg
165 170 175

Glu Phe Ser Phe Val Gln Ser Ser Leu Gly Phe Val Ala Leu Val Leu
180 185 190

Ser Thr Leu His Thr Leu Thr Tyr Gly Trp Thr Arg Ala Phe Glu Glu
195 200 205

Ser Arg Tyr Lys Phe Tyr Leu Pro Pro Thr Phe Thr Leu Thr Leu Leu
210 215 220

Val Pro Cys Val Val Ile Leu Ala Lys Ala Leu Phe Leu
225 230 235

<210> 16
<211> 146
<212> PRT
<213> Homo sapiens

<400> 16

Ala Arg Gly Gly Arg Leu Arg Trp Arg Arg Leu Asp Asp Cys Leu Ser
1 5 10 15

Ala Ala Glu Ser Asp Thr Val Ala Tyr Glu Asp Leu Ser Glu Asp Tyr
20 25 30

Thr Gln Lys Lys Trp Lys Gly Leu Ala Leu Ser Gln Arg Ala Leu His
35 40 45

Trp Asn Met Met Leu Glu Asn Asp Arg Ser Met Ala Ser Leu Ala Gly
50 55 60

Arg Asn Met Met Glu Ser Ser Glu Leu Thr Pro Lys Gln Glu Ile Phe
65 70 75 80

Lys Gly Ser Glu Ser Ser Asn Ser Thr Ser Gly Gly Leu Phe Gly Val
85 90 95

Val Pro Gly Gly Thr Glu Thr Gly Asp Val Cys Glu Asp Thr Phe Lys
100 105 110

Glu Leu Glu Gly Gln Pro Ser Asn Glu Glu Gly Ser Arg Leu Glu Ser
115 120 125

Asp Phe Leu Glu Ile Ile Asp Glu Asp Lys Lys Lys Ser Thr Lys Asp
130 135 140

Arg Tyr
145

<210> 17
<211> 364
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(364)
<223> n = A, T, G, or C

<400> 17
cggacgcgtg ggctcggggc aggtttccgg gagcagatgc caaaaagact ttttcataga 60

gaagaggctt tcttttgtag agacagaata aaaataattg gtatgtttct gtttgttccc 120

tccccctccc ccttgtgtga taccacatgt gtatagtatt taagtgaaac tcaagccctc 180

aaggcccaac ttctctgtct atattgtaat atagaatttc gaagagacat tttcactttt 240

tacacattgg gcacaaagat aagctttgat taaagtagta agtaaaaggc tacctaggaa 300

atacttcagt gaattctaac nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnngaaa 360

cagg 364

<210> 18
<211> 5773
<212> DNA
<213> Homo sapiens

<400> 18

```
tttcgtgatc atagctgggg gaggctgagc gtgggagcgg tgctgccagt cctgcctgaa      60
aacgcgaaat gagtcttgct tggttctccc tccactgggc gtgagagccc ctgcccagga     120
ggcccaggac aaatggcccc atagtggaaa ctgggaagct tttaggcatc tgatcagagc     180
gggagccagc cgggggacca cagtgctgga caggccaacc aactcaaact tgaagacatg     240
aaatccccaa ggagaaccac tttgtgcctc atgtttattg tgatttattc ttccaaagct     300
gcactgaact ggaattacga gtctactatt catcctttga gtcttcatga acatgaacca     360
gctggtgaag aggcactgag gcaaaaacga gccgttgcca caaaaagtcc tacggctgaa     420
gaatacactg ttaatattga gatcagtttt gaaaatgcat ccttcctgga tcctatcaaa     480
gcctacttga acagcctcag ttttccaatt catgggaata acactgacca aattactgac     540
attttgagca taaatgtgac aacagtctgc agacctgctg gaaatgaaat ctggtgctcc     600
tgcgagacag gttatgggtg gcctcgggaa aggtgtcttc acaatctcat ttgtcaagag     660
cgtgacgtct tcctcccagg gcaccattgc agttgcctta aagaactgcc tcccaatgga     720
ccttttttgcc tgcttcagga agatgttacc ctgaacatga gagtcagact aaatgtaggc     780
tttcaagaag acctcatgaa cacttcctcc gccctctata ggtcctacaa gaccgacttg     840
gaaacagcgt tccggaaggg ttacggaatt ttaccaggct tcaagggcgt gactgtgaca     900
gggttcaagt ctggaagtgt ggttgtgaca tatgaagtca agactacacc accatcactt     960
gagttaatac ataaagccaa tgaacaagtt gtacagagcc tcaatcagac ctacaaaatg    1020
gactacaact cctttcaagc agttactatc aatgaaagca atttctttgt cacaccagaa    1080
atcatctttg aaggggacac agtcagtctg gtgtgtgaaa aggaagtttt gtcctccaat    1140
gtgtcttggc gctatgaaga acagcagttg gaaatccaga acagcagcag attctcgatt    1200
tacaccgcac ttttcaacaa catgacttcg gtgtccaagc tcaccatcca caacatcact    1260
ccaggtgatg caggtgaata tgtttgcaaa ctgatattag acatttttga atatgagtgc    1320
aagaagaaaa tagatgttat gcccatccaa attttggcaa atgaagaaat gaaggtgatg    1380
tgcgacaaca atcctgtatc tttgaactgc tgcagtcagg gtaatgttaa ttggagcaaa    1440
gtagaatgga agcaggaagg aaaaataaat attccaggaa cccctgagac agacatagat    1500
tctagctgca gcagatacac cctcaaggct gatggaaccc agtgcccaag cgggtcgtct    1560
ggaacaacag tcatctacac ttgtgagttc atcagtgcct atggagccag aggcagtgca    1620
aacataaaag tgacattcat ctctgtggcc aatctaacaa taacccccgga cccaatttct    1680
gtttctgagg gacaaaactt ttctataaaa tgcatcagtg atgtgagtaa ctatgatgag    1740
gtttattgga acacttctgc tggaattaaa atataccaaa gattttatac cacgaggagg    1800
tatcttgatg gagcagaatc agtactgaca gtcaagacct cgaccaggga gtggaatgga    1860
acctatcact gcatatttag atataagaat tcatacagta ttgcaaccaa agacgtcatt    1920
```

gttcacccgc tgcctctaaa gctgaacatc atggttgatc ctttggaagc tactgtttca 1980
tgcagtggtt cccatcacat caagtgctgc atagaggagg atggagacta caaagttact 2040
ttccatatgg gttcctcatc ccttcctgct gcaaaagaag ttaacaaaaa acaagtgtgc 2100
tacaaacaca atttcaatgc aagctcagtt tcctggtgtt caaaaactgt tgatgtgtgt 2160
tgtcacttta ccaatgctgc taataattca gtttggagcc catctatgaa gctgaatctg 2220
gttcctgggg aaaacatcac atgccaggat cccgtaatag gtgtcggaga gccggggaaa 2280
gtcatccaga agctatgccg gttctcaaac gttcccagca gccctgagga gtcccattag 2340
gcgggaccat cacttacaaa tgtgtaggct cccagtgggg ggtagaagag aaatgactgc 2400
atctctgccc caataaacag tctgctccag atggctaagg ctttgatcaa gagcccctct 2460
caggatgaga tgctccctac atacctgaag gatctttcta ttagcataga caaagcggaa 2520
catgaaatca gctcttctcc tgggagtctg ggagccatta ttaacatcct tgatctgctc 2580
tcaacagttc caacccaagt aaattcagaa atgatgacgc acgtgctctc tacggttaat 2640
gtcatccttg gcaagcccgt cttgaacacc tggaaggttt tacaacagca atggaccaat 2700
cagagttcac agctactaca ttcagtggaa agattttccc aagcattaca gtcaggagat 2760
agccctcctt tgtccttctc ccaaactaat gtgcagatga gcagcacggt aatcaagtcc 2820
agccacccag aaacctatca acagaggttt gttttcccat actttgacct ctggggcaat 2880
gtggtcattg acaagagcta cctagaaaac ttgcagtcgg attcgtctat tgtcaccatg 2940
gctttcccaa ctctccaagc catccttgct caggatatcc aggaaaataa ctttgcagag 3000
agcttagtga tgacaaccac tgtcagccac aatacgacta tgccattcag gatttcaatg 3060
acttttaaga acaatagccc ttcaggcggc gaaacgaagt gtgtcttctg gaacttcagg 3120
cttgccaaca acacaggggg gtgggacagc agtgggtgct atgttgaaga aggtgatggg 3180
gacaatgtca cctgtatctg tgaccaccta acatcattct ccatcctcat gtcccctgac 3240
tccccagatc ctagttctct cctgggaata ctcctggata ttatttctta tgttggggtg 3300
ggcttttcca tcttgagctt ggcagcctgt ctagttgtgg aagctgtggt gtggaaatcg 3360
gtgaccaaga atcggacttc ttatatgcgc cacacctgca tagtgaatat cgctgcctcc 3420
cttctgggtc gccaacacct ggttcattgg ggtcgctgcc atccaggaca atcgctacat 3480
actctgcaag acagcctgtg tggctgccac cttcttcatc cacttcttct acctcagcgt 3540
cttcttctgg atgctgacac tgggcctcat gctgttctat cgcctggttt tcattctgca 3600
tgaaacaagc aggtccactc agaaagccat tgccttctgt cttggctatg gctgcccact 3660
tgccatctcg gtcatcacgc tgggagccac ccagccccgg gaagtctata cgaggaagaa 3720
tgtctgttgg ctcaactggg aggacaccaa ggccctgctg gctttcgcca tcccagcact 3780

gatcattgtg gtggtgaaca taaccatcac tattgtggtc atcaccaaga tcctgaggcc 3840

ttccattgga gacaagccat gcaagcagga gaagagcagc ctgtttcaga tcagcaagag 3900

cattggggtc ctcacaccac tcttgggcct cacttggggt tttggtctca ccactgtgtt 3960

cccagggacc aaccttgtgt tccatatcat atttgccatc ctcaatgtct tccagggatt 4020

attcatttta ctctttggat gcctctggga tctgaaggta caggaagctt tgctgaataa 4080

gttttcattg tcgagatggt cttcacagca ctcaaagtca acatccctgg gttcatccac 4140

acctgtgttt tctatgagtt ctccaatatc aaggagattt aacaatttgt ttggtaaaac 4200

aggaacgtat aatgtttcca ccccagaagc aaccagctca tccctggaaa actcatccag 4260

tgcttcttcg ttgctcaact aagaacagga taatccaacc tacgtgacct cccggggaca 4320

gtggctgtgc ttttaaaaag agatgcttgc caagcaatgg ggaacgtgtt ctcggggcag 4380

gtttccggga gcagatgcca aaaagacttt ttcatagaga agaggctttc ttttgtaaag 4440

acagaataaa aataattgtt atgtttctgt ttgttccctc cccctccccc ttgtgtgata 4500

ccacatgtgt atagtattta agtgaaactc aagccctcaa ggcccaactt ctctgtctat 4560

attgtaatat agaatttcga agagacattt tcacttttta cacattgggc acaaagataa 4620

gctttgatta aagtagtaag taaaaggcta cctaggaaat acttcagtga attctaagaa 4680

ggaaggaagg aagaaaggaa ggaaagaagg gagggaaaca gggagaaagg gaaaaagaag 4740

aaaaagagaa agatgaaaat aggaacaaat aaagacaaac aacattaagg gccatattgt 4800

aagatttcca tgttaatgat ctaatataat cactcagtgc aacattgaga attttttttt 4860

aatggctcaa aaatggaaac tgaaagcaag tcatggggaa tgaatacttt gggcagtatc 4920

ttcctcatgt cttcttagct aagaggagga aaaaaaggct gaaaaaatag ggaggaaatt 4980

ccttcatcag aacgacttca agtggataac aatatttata agaaatgaat ggaaggaaat 5040

atgatcctcc tgagactaac tttgtatgtt aaggtttgaa ctaagtgaat gtatctgcag 5100

aggaagtatt acaaagatat gtcattagat ccaagtgctg attaaatttt tatagtttat 5160

cagaaaagcc ttatatttta gtttgttcca cattttgaaa gcaaaaaata tatatttgat 5220

ataccettca attgccaaat ttgatatgtt gcactgaaga cagaccctgt catatattta 5280

atggcttcaa gcaggtactt ctctgtgcat tatagaatag attttaataa tcttatagca 5340

ttgtatatta ttattgctgt tgtcactgtt attattattg tggatactgg cccttggtgt 5400

gttgcatagc tccctatgta ttctctgttt ccatctttaa gttcccagac caatatacat 5460

taagagtttt gcatggtcta aattgtgttt attccaacca cgtggaaagc tcctggaaag 5520

aaattttaca ttcggttgtt ctgtgctcct aatgacactt gaccttgttg aacaaatggc 5580

agagcctttc ccaaggattt gattgtttgt gaattatctg catgtgtgct ttttttggt 5640

gtgtatgtat ttcattaaaa aatataaata cttatgaaaa ttgcacgcat attagagtta 5700

accatgtact attgatacag caacgctaca ttgcaaataa aagtccgatc ccaaaaggag 5760

aatgagacaa aaa 5773

<210> 19
<211> 5714
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (272)..(4312)

<400> 19
ctgcacgacc ggtccggaat tcccgggtcg acgatttcgt gatcatagct gggggaggct 60

gagcgtggga gcggtgctgc cagtcctgcc tgaaaacgcg aaatgagtct tgcttggttc 120

tccctccact gggcgtgaga gcccctgccc aggaggccca ggacaaatgg ccccatagtg 180

gaaactggga agcttttagg catctgatca gagcgggagc cagccggggg accacagtgc 240

tggacaggcc aaccaactca aacttgaaga c atg aaa tcc cca agg aga acc 292
Met Lys Ser Pro Arg Arg Thr
1 5

act ttg tgc ctc atg ttt att gtg att tat tct tcc aaa gct gca ctg 340
Thr Leu Cys Leu Met Phe Ile Val Ile Tyr Ser Ser Lys Ala Ala Leu
10 15 20

aac tgg aat tac gag tct act att cat cct ttg agt ctt cat gaa cat 388
Asn Trp Asn Tyr Glu Ser Thr Ile His Pro Leu Ser Leu His Glu His
25 30 35

gaa cca gct ggt gaa gag gca ctg agg caa aaa cga gcc gtt gcc aca 436
Glu Pro Ala Gly Glu Glu Ala Leu Arg Gln Lys Arg Ala Val Ala Thr
40 45 50 55

aaa agt cct acg gct gaa gaa tac act gtt aat att gag atc agt ttt 484
Lys Ser Pro Thr Ala Glu Glu Tyr Thr Val Asn Ile Glu Ile Ser Phe
60 65 70

gaa aat gca tcc ttc ctg gat cct atc aaa gcc tac ttg aac agc ctc 532
Glu Asn Ala Ser Phe Leu Asp Pro Ile Lys Ala Tyr Leu Asn Ser Leu
75 80 85

agt ttt cca att cat ggg aat aac act gac caa att act gac att ttg 580
Ser Phe Pro Ile His Gly Asn Asn Thr Asp Gln Ile Thr Asp Ile Leu
90 95 100

agc ata aat gtg aca aca gtc tgc aga cct gct gga aat gaa atc tgg 628
Ser Ile Asn Val Thr Thr Val Cys Arg Pro Ala Gly Asn Glu Ile Trp
105 110 115

tgc tcc tgc gag aca ggt tat ggg tgg cct cgg gaa agg tgt ctt cac 676
Cys Ser Cys Glu Thr Gly Tyr Gly Trp Pro Arg Glu Arg Cys Leu His
120 125 130 135

aat ctc att tgt caa gag cgt gac gtc ttc ctc cca ggg cac cat tgc 724
Asn Leu Ile Cys Gln Glu Arg Asp Val Phe Leu Pro Gly His His Cys
140 145 150

agt tgc ctt aaa gaa ctg cct ccc aat gga cct ttt tgc ctg ctt cag 772
Ser Cys Leu Lys Glu Leu Pro Pro Asn Gly Pro Phe Cys Leu Leu Gln
155 160 165

gaa gat gtt acc ctg aac atg aga gtc aga cta aat gta ggc ttt caa 820
Glu Asp Val Thr Leu Asn Met Arg Val Arg Leu Asn Val Gly Phe Gln
170 175 180

gaa gac ctc atg aac act tcc tcc gcc ctc tat agg tcc tac aag acc 868
Glu Asp Leu Met Asn Thr Ser Ser Ala Leu Tyr Arg Ser Tyr Lys Thr
185 190 195

gac ttg gaa aca gcg ttc cgg aag ggt tac gga att tta cca ggc ttc 916
Asp Leu Glu Thr Ala Phe Arg Lys Gly Tyr Gly Ile Leu Pro Gly Phe
200 205 210 215

aag ggc gtg act gtg aca ggg ttc aag tct gga agt gtg gtt gtg aca 964
Lys Gly Val Thr Val Thr Gly Phe Lys Ser Gly Ser Val Val Val Thr
220 225 230

tat gaa gtc aag act aca cca cca tca ctt gag tta ata cat aaa gcc 1012
Tyr Glu Val Lys Thr Thr Pro Pro Ser Leu Glu Leu Ile His Lys Ala
235 240 245

aat gaa caa gtt gta cag agc ctc aat cag acc tac aaa atg gac tac 1060
Asn Glu Gln Val Val Gln Ser Leu Asn Gln Thr Tyr Lys Met Asp Tyr
250 255 260

aac tcc ttt caa gca gtt act atc aat gaa agc aat ttc ttt gtc aca 1108
Asn Ser Phe Gln Ala Val Thr Ile Asn Glu Ser Asn Phe Phe Val Thr
265 270 275

cca gaa atc atc ttt gaa ggg gac aca gtc agt ctg gtg tgt gaa aag 1156
Pro Glu Ile Ile Phe Glu Gly Asp Thr Val Ser Leu Val Cys Glu Lys
280 285 290 295

gaa gtt ttg tcc tcc aat gtg tct tgg cgc tat gaa gaa cag cag ttg 1204
Glu Val Leu Ser Ser Asn Val Ser Trp Arg Tyr Glu Glu Gln Gln Leu
300 305 310

gaa atc cag aac agc agc aga ttc tcg att tac acc gca ctt ttc aac 1252
Glu Ile Gln Asn Ser Ser Arg Phe Ser Ile Tyr Thr Ala Leu Phe Asn
315 320 325

aac atg act tcg gtg tcc aag ctc acc atc cac aac atc act cca ggt 1300
Asn Met Thr Ser Val Ser Lys Leu Thr Ile His Asn Ile Thr Pro Gly
330 335 340

gat gca ggt gaa tat gtt tgc aaa ctg ata tta gac att ttt gaa tat 1348
Asp Ala Gly Glu Tyr Val Cys Lys Leu Ile Leu Asp Ile Phe Glu Tyr
345 350 355

gag tgc aag aag aaa ata gat gtt atg ccc atc caa att ttg gca aat 1396
Glu Cys Lys Lys Lys Ile Asp Val Met Pro Ile Gln Ile Leu Ala Asn
360 365 370 375

gaa gaa atg aag gtg atg tgc gac aac aat cct gta tct ttg aac tgc 1444
Glu Glu Met Lys Val Met Cys Asp Asn Asn Pro Val Ser Leu Asn Cys
380 385 390

tgc agt cag ggt aat gtt aat tgg agc aaa gta gaa tgg aag cag gaa 1492
Cys Ser Gln Gly Asn Val Asn Trp Ser Lys Val Glu Trp Lys Gln Glu

395 400 405

gga aaa ata aat att cca gga acc cct gag aca gac ata gat tct agc 1540
Gly Lys Ile Asn Ile Pro Gly Thr Pro Glu Thr Asp Ile Asp Ser Ser
410 415 420

tgc agc aga tac acc ctc aag gct gat gga acc cag tgc cca agc ggg 1588
Cys Ser Arg Tyr Thr Leu Lys Ala Asp Gly Thr Gln Cys Pro Ser Gly
425 430 435

tcg tct gga aca aca gtc atc tac act tgt gag ttc atc agt gcc tat 1636
Ser Ser Gly Thr Thr Val Ile Tyr Thr Cys Glu Phe Ile Ser Ala Tyr
440 445 450 455

gga gcc aga ggc agt gca aac ata aaa gtg aca ttc atc tct gtg gcc 1684
Gly Ala Arg Gly Ser Ala Asn Ile Lys Val Thr Phe Ile Ser Val Ala
460 465 470

aat cta aca ata acc ccg gac cca att tct gtt tct gag gga caa aac 1732
Asn Leu Thr Ile Thr Pro Asp Pro Ile Ser Val Ser Glu Gly Gln Asn
475 480 485

ttt tct ata aaa tgc atc agt gat gtg agt aac tat gat gag gtt tat 1780
Phe Ser Ile Lys Cys Ile Ser Asp Val Ser Asn Tyr Asp Glu Val Tyr
490 495 500

tgg aac act tct gct gga att aaa ata tac caa aga ttt tat acc acg 1828
Trp Asn Thr Ser Ala Gly Ile Lys Ile Tyr Gln Arg Phe Tyr Thr Thr
505 510 515

agg agg tat ctt gat gga gca gaa tca gta ctg aca gtc aag acc tcg 1876
Arg Arg Tyr Leu Asp Gly Ala Glu Ser Val Leu Thr Val Lys Thr Ser
520 525 530 535

acc agg gag tgg aat gga acc tat cac tgc ata ttt aga tat aag aat 1924
Thr Arg Glu Trp Asn Gly Thr Tyr His Cys Ile Phe Arg Tyr Lys Asn
540 545 550

tca tac agt att gca acc aaa gac gtc att gtt cac ccg ctg cct cta 1972
Ser Tyr Ser Ile Ala Thr Lys Asp Val Ile Val His Pro Leu Pro Leu
555 560 565

aag ctg aac atc atg gtt gat cct ttg gaa gct act gtt tca tgc agt 2020
Lys Leu Asn Ile Met Val Asp Pro Leu Glu Ala Thr Val Ser Cys Ser
570 575 580

ggt tcc cat cac atc aag tgc tgc ata gag gag gat gga gac tac aaa 2068
Gly Ser His His Ile Lys Cys Cys Ile Glu Glu Asp Gly Asp Tyr Lys
585 590 595

gtt act ttc cat atg ggt tcc tca tcc ctt cct gct gca aaa gaa gtt 2116
Val Thr Phe His Met Gly Ser Ser Ser Leu Pro Ala Ala Lys Glu Val
600 605 610 615

aac aaa aaa caa gtg tgc tac aaa cac aat ttc aat gca agc tca gtt 2164
Asn Lys Lys Gln Val Cys Tyr Lys His Asn Phe Asn Ala Ser Ser Val
620 625 630

tcc tgg tgt tca aaa act gtt gat gtg tgt tgt cac ttt acc aat gct 2212
Ser Trp Cys Ser Lys Thr Val Asp Val Cys Cys His Phe Thr Asn Ala
635 640 645

gct aat aat tca gtt tgg agc cca tct atg aag ctg aat ctg gtt cct 2260

Ala Asn Asn Ser Val Trp Ser Pro Ser Met Lys Leu Asn Leu Val Pro
650 655 660

ggg gaa aac atc aca tgc cag gat ccc gta ata ggt gtc gga gag ccg 2308
Gly Glu Asn Ile Thr Cys Gln Asp Pro Val Ile Gly Val Gly Glu Pro
665 670 675

ggg aaa gtc atc cag aag cta tgc cgg ttc tca aac gtt ccc agc agc 2356
Gly Lys Val Ile Gln Lys Leu Cys Arg Phe Ser Asn Val Pro Ser Ser
680 685 690 695

cct gag agt ccc att ggc ggg acc atc act tac aaa tgt gta ggc tcc 2404
Pro Glu Ser Pro Ile Gly Gly Thr Ile Thr Tyr Lys Cys Val Gly Ser
700 705 710

cag tgg gag gag aag aga aat gac tgc atc tct gcc cca ata aac agt 2452
Gln Trp Glu Glu Lys Arg Asn Asp Cys Ile Ser Ala Pro Ile Asn Ser
715 720 725

ctg ctc cag atg gct aag gct ttg atc aag agc ccc tct cag gat gag 2500
Leu Leu Gln Met Ala Lys Ala Leu Ile Lys Ser Pro Ser Gln Asp Glu
730 735 740

atg ctc cct aca tac ctg aag gat ctt tct att agc ata gac aaa gcg 2548
Met Leu Pro Thr Tyr Leu Lys Asp Leu Ser Ile Ser Ile Asp Lys Ala
745 750 755

gaa cat gaa atc agc tct tct cct ggg agt ctg gga gcc att att aac 2596
Glu His Glu Ile Ser Ser Ser Pro Gly Ser Leu Gly Ala Ile Ile Asn
760 765 770 775

atc ctt gat ctg ctc tca aca gtt cca acc caa gta aat tca gaa atg 2644
Ile Leu Asp Leu Leu Ser Thr Val Pro Thr Gln Val Asn Ser Glu Met
780 785 790

atg acg cac gtg ctc tct acg gtt aat gtc atc ctt ggc aag ccc gtc 2692
Met Thr His Val Leu Ser Thr Val Asn Val Ile Leu Gly Lys Pro Val
795 800 805

ttg aac acc tgg aag gtt tta caa cag caa tgg acc aat cag agt tca 2740
Leu Asn Thr Trp Lys Val Leu Gln Gln Gln Trp Thr Asn Gln Ser Ser
810 815 820

cag cta cta cat tca gtg gaa aga ttt tcc caa gca tta cag tca gga 2788
Gln Leu Leu His Ser Val Glu Arg Phe Ser Gln Ala Leu Gln Ser Gly
825 830 835

gat agc cct cct ttg tcc ttc tcc caa act aat gtg cag atg agc agc 2836
Asp Ser Pro Pro Leu Ser Phe Ser Gln Thr Asn Val Gln Met Ser Ser
840 845 850 855

acg gta atc aag tcc agc cac cca gaa acc tat caa cag agg ttt gtt 2884
Thr Val Ile Lys Ser Ser His Pro Glu Thr Tyr Gln Gln Arg Phe Val
860 865 870

ttc cca tac ttt gac ctc tgg ggc aat gtg gtc att gac aag agc tac 2932
Phe Pro Tyr Phe Asp Leu Trp Gly Asn Val Val Ile Asp Lys Ser Tyr
875 880 885

cta gaa aac ttg cag tcg gat tcg tct att gtc acc atg gct ttc cca 2980
Leu Glu Asn Leu Gln Ser Asp Ser Ser Ile Val Thr Met Ala Phe Pro
890 895 900

act ctc caa gcc atc ctt gct cag gat atc cag gaa aat aac ttt gca 3028
Thr Leu Gln Ala Ile Leu Ala Gln Asp Ile Gln Glu Asn Asn Phe Ala
905 910 915

gag agc tta gtg atg aca acc act gtc agc cac aat acg act atg cca 3076
Glu Ser Leu Val Met Thr Thr Thr Val Ser His Asn Thr Thr Met Pro
920 925 930 935

ttc agg att tca atg act ttt aag aac aat agc cct tca ggc ggc gaa 3124
Phe Arg Ile Ser Met Thr Phe Lys Asn Asn Ser Pro Ser Gly Gly Glu
940 945 950

acg aag tgt gtc ttc tgg aac ttc agg ctt gcc aac aac aca ggg ggg 3172
Thr Lys Cys Val Phe Trp Asn Phe Arg Leu Ala Asn Asn Thr Gly Gly
955 960 965

tgg gac agc agt ggg tgc tat gtt gaa gaa ggt gat ggg gac aat gtc 3220
Trp Asp Ser Ser Gly Cys Tyr Val Glu Glu Gly Asp Gly Asp Asn Val
970 975 980

acc tgt atc tgt gac cac cta aca tca ttc tcc atc ctc atg tcc cct 3268
Thr Cys Ile Cys Asp His Leu Thr Ser Phe Ser Ile Leu Met Ser Pro
985 990 995

gac tcc cca gat cct agt tct ctc ctg gga ata ctc ctg gat att 3313
Asp Ser Pro Asp Pro Ser Ser Leu Leu Gly Ile Leu Leu Asp Ile
1000 1005 1010

att tct tat gtt ggg gtg ggc ttt tcc atc ttg agc ttg gca gcc 3358
Ile Ser Tyr Val Gly Val Gly Phe Ser Ile Leu Ser Leu Ala Ala
1015 1020 1025

tgt cta gtt gtg gaa gct gtg gtg tgg aaa tcg gtg acc aag aat 3403
Cys Leu Val Val Glu Ala Val Val Trp Lys Ser Val Thr Lys Asn
1030 1035 1040

cgg act tct tat atg cgc cac acc tgc ata gtg aat atc gct gcc 3448
Arg Thr Ser Tyr Met Arg His Thr Cys Ile Val Asn Ile Ala Ala
1045 1050 1055

tcc ctt ctg gtc gcc aac acc tgg ttc att gtg gtc gct gcc atc 3493
Ser Leu Leu Val Ala Asn Thr Trp Phe Ile Val Val Ala Ala Ile
1060 1065 1070

cag gac aat cgc tac ata ctc tgc aag aca gcc tgt gtg gct gcc 3538
Gln Asp Asn Arg Tyr Ile Leu Cys Lys Thr Ala Cys Val Ala Ala
1075 1080 1085

acc ttc ttc atc cac ttc ttc tac ctc agc gtc ttc ttc tgg atg 3583
Thr Phe Phe Ile His Phe Phe Tyr Leu Ser Val Phe Phe Trp Met
1090 1095 1100

ctg aca ctg ggc ctc atg ctg ttc tat cgc ctg gtt ttc att ctg 3628
Leu Thr Leu Gly Leu Met Leu Phe Tyr Arg Leu Val Phe Ile Leu
1105 1110 1115

cat gaa aca agc agg tcc act cag aaa gcc att gcc ttc tgt ctt 3673
His Glu Thr Ser Arg Ser Thr Gln Lys Ala Ile Ala Phe Cys Leu
1120 1125 1130

ggc tat ggc tgc cca ctt gcc atc tcg gtc atc acg ctg gga gcc 3718
Gly Tyr Gly Cys Pro Leu Ala Ile Ser Val Ile Thr Leu Gly Ala
1135 1140 1145

```
acc  cag ccc cgg gaa gtc  tat acg agg aag aat  gtc tgt tgg ctc       3763
Thr  Gln Pro Arg Glu Val  Tyr Thr Arg Lys Asn  Val Cys Trp Leu
1150                 1155                 1160

aac  tgg gag gac acc aag  gcc ctg ctg gct ttc  gcc atc cca gca       3808
Asn  Trp Glu Asp Thr Lys  Ala Leu Leu Ala Phe  Ala Ile Pro Ala
1165                 1170                 1175

ctg  atc att gtg gtg gtg  aac ata acc atc act  att gtg gtc atc       3853
Leu  Ile Ile Val Val Val  Asn Ile Thr Ile Thr  Ile Val Val Ile
1180                 1185                 1190

acc  aag atc ctg agg cct  tcc att gga gac aag  cca tgc aag cag       3898
Thr  Lys Ile Leu Arg Pro  Ser Ile Gly Asp Lys  Pro Cys Lys Gln
1195                 1200                 1205

gag  aag agc agc ctg ttt  cag atc agc aag agc  att ggg gtc ctc       3943
Glu  Lys Ser Ser Leu Phe  Gln Ile Ser Lys Ser  Ile Gly Val Leu
1210                 1215                 1220

aca  cca ctc ttg ggc ctc  act tgg ggt ttt ggt  ctc acc act gtg       3988
Thr  Pro Leu Leu Gly Leu  Thr Trp Gly Phe Gly  Leu Thr Thr Val
1225                 1230                 1235

ttc  cca ggg acc aac ctt  gtg ttc cat atc ata  ttt gcc atc ctc       4033
Phe  Pro Gly Thr Asn Leu  Val Phe His Ile Ile  Phe Ala Ile Leu
1240                 1245                 1250

aat  gtc ttc cag gga tta  ttc att tta ctc ttt  gga tgc ctc tgg       4078
Asn  Val Phe Gln Gly Leu  Phe Ile Leu Leu Phe  Gly Cys Leu Trp
1255                 1260                 1265

gat  ctg aag gta cag gaa  gct ttg ctg aat aag  ttt tca ttg tcg       4123
Asp  Leu Lys Val Gln Glu  Ala Leu Leu Asn Lys  Phe Ser Leu Ser
1270                 1275                 1280

aga  tgg tct tca cag cac  tca aag tca aca tcc  ctg ggt tca tcc       4168
Arg  Trp Ser Ser Gln His  Ser Lys Ser Thr Ser  Leu Gly Ser Ser
1285                 1290                 1295

aca  cct gtg ttt tct atg  agt tct cca ata tca  agg aga ttt aac       4213
Thr  Pro Val Phe Ser Met  Ser Ser Pro Ile Ser  Arg Arg Phe Asn
1300                 1305                 1310

aat  ttg ttt ggt aaa aca  gga acg tat aat gtt  tcc acc cca gaa       4258
Asn  Leu Phe Gly Lys Thr  Gly Thr Tyr Asn Val  Ser Thr Pro Glu
1315                 1320                 1325

gca  acc agc tca tcc ctg  gaa aac tca tcc agt  gct tct tcg ttg       4303
Ala  Thr Ser Ser Ser Leu  Glu Asn Ser Ser Ser  Ala Ser Ser Leu
1330                 1335                 1340

ctc  aac taa gaacaggata atccaaccta cgtgacctcc cggggacagt             4352
Leu  Asn
1345

ggctgtgctt ttaaaaagag atgcttgcaa agcaatgggg aacgtgttct cggggcaggt   4412

ttccgggagc agatgccaaa aagacttttt catagagaag aggctttctt ttgtaaagac   4472

agaataaaaa taattgttat gtttctgttt gttccctccc cctccccctt gtgtgatacc   4532
```

acatgtgtat agtatttaag tgaaactcaa gccctcaagg cccaacttct ctgtctatat 4592

tgtaatatag aatttcgaag agacattttc actttttaca cattgggcac aaagataagc 4652

tttgattaaa gtagtaagta aaaggctacc taggaaatac ttcagtgaat tctaagaagg 4712

aaggaaggaa gaaaggaagg aaagaaggga gggaaacagg gagaaaggga aaaagaagaa 4772

aaagagaaag atgaaaatag gaacaaataa agacaaacaa cattaagggc catattgtaa 4832

gatttccatg ttaatgatct aatataatca ctcagtgcaa cattgagaat ttttttttaa 4892

tggctcaaaa atggaaactg aaagcaagtc atggggaatg aatactttgg gcagtatctt 4952

cctcatgtct tcttagctaa gaggaggaaa aaaaggctga aaaaataggg aggaaattcc 5012

ttcatcagaa cgacttcaag tggataacaa tatttataag aaatgaatgg aaggaaatat 5072

gatcctcctg agactaactt tgtatgttaa ggtttgaact aagtgaatgt atctgcagag 5132

gaagtattac aaagatatgt cattagatcc aagtgctgat taaattttta tagtttatca 5192

gaaaagcctt atattttagt ttgttccaca ttttgaaagc aaaaaatata tatttgatat 5252

acccttcaat tgccaaattt gatatgttgc actgaagaca gaccctgtca tatatttaat 5312

ggcttcaagc aggtacttct ctgtgcatta tagaatagat tttaataatc ttatagcatt 5372

gtatattatt attgctgttg tcactgttat tattattgtg gatactggcc cttggtgtgt 5432

tgcatagctc cctatgtatt ctctgtttcc atctttaagt tcccagacca atatacatta 5492

agagttttgc atggtctaaa ttgtgtttat tccaaccacg tggaaagctc ctggaaagaa 5552

attttacatt cggttgttct gtgctcctaa tgacacttga ccttgttgaa caaatggcag 5612

agcctttccc aaggatttga ttgtttgtga attatctgca tgtgtgcttt tttttggtgt 5672

gtatttcatt aaaaaatata aatatttatg aaaaaaaaaa aa 5714

<210> 20
<211> 1346
<212> PRT
<213> Homo sapiens

<400> 20

Met Lys Ser Pro Arg Arg Thr Thr Leu Cys Leu Met Phe Ile Val Ile
1 5 10 15

Tyr Ser Ser Lys Ala Ala Leu Asn Trp Asn Tyr Glu Ser Thr Ile His
20 25 30

Pro Leu Ser Leu His Glu His Glu Pro Ala Gly Glu Glu Ala Leu Arg
35 40 45

Gln Lys Arg Ala Val Ala Thr Lys Ser Pro Thr Ala Glu Glu Tyr Thr
50 55 60

Val Asn Ile Glu Ile Ser Phe Glu Asn Ala Ser Phe Leu Asp Pro Ile
65 70 75 80

Lys Ala Tyr Leu Asn Ser Leu Ser Phe Pro Ile His Gly Asn Asn Thr
85 90 95

Asp Gln Ile Thr Asp Ile Leu Ser Ile Asn Val Thr Thr Val Cys Arg
100 105 110

Pro Ala Gly Asn Glu Ile Trp Cys Ser Cys Glu Thr Gly Tyr Gly Trp
115 120 125

Pro Arg Glu Arg Cys Leu His Asn Leu Ile Cys Gln Glu Arg Asp Val
130 135 140

Phe Leu Pro Gly His His Cys Ser Cys Leu Lys Glu Leu Pro Pro Asn
145 150 155 160

Gly Pro Phe Cys Leu Leu Gln Glu Asp Val Thr Leu Asn Met Arg Val
165 170 175

Arg Leu Asn Val Gly Phe Gln Glu Asp Leu Met Asn Thr Ser Ser Ala
180 185 190

Leu Tyr Arg Ser Tyr Lys Thr Asp Leu Glu Thr Ala Phe Arg Lys Gly
195 200 205

Tyr Gly Ile Leu Pro Gly Phe Lys Gly Val Thr Val Thr Gly Phe Lys
210 215 220

Ser Gly Ser Val Val Val Thr Tyr Glu Val Lys Thr Thr Pro Pro Ser
225 230 235 240

Leu Glu Leu Ile His Lys Ala Asn Glu Gln Val Val Gln Ser Leu Asn
245 250 255

Gln Thr Tyr Lys Met Asp Tyr Asn Ser Phe Gln Ala Val Thr Ile Asn
260 265 270

Glu Ser Asn Phe Phe Val Thr Pro Glu Ile Ile Phe Glu Gly Asp Thr
275 280 285

Val Ser Leu Val Cys Glu Lys Glu Val Leu Ser Ser Asn Val Ser Trp
290 295 300

Arg Tyr Glu Glu Gln Gln Leu Glu Ile Gln Asn Ser Ser Arg Phe Ser
305 310 315 320

```
Ile Tyr Thr Ala Leu Phe Asn Asn Met Thr Ser Val Ser Lys Leu Thr
                325                 330                 335

Ile His Asn Ile Thr Pro Gly Asp Ala Gly Glu Tyr Val Cys Lys Leu
            340                 345                 350

Ile Leu Asp Ile Phe Glu Tyr Glu Cys Lys Lys Lys Ile Asp Val Met
        355                 360                 365

Pro Ile Gln Ile Leu Ala Asn Glu Glu Met Lys Val Met Cys Asp Asn
    370                 375                 380

Asn Pro Val Ser Leu Asn Cys Cys Ser Gln Gly Asn Val Asn Trp Ser
385                 390                 395                 400

Lys Val Glu Trp Lys Gln Glu Gly Lys Ile Asn Ile Pro Gly Thr Pro
                405                 410                 415

Glu Thr Asp Ile Asp Ser Ser Cys Ser Arg Tyr Thr Leu Lys Ala Asp
            420                 425                 430

Gly Thr Gln Cys Pro Ser Gly Ser Ser Gly Thr Thr Val Ile Tyr Thr
        435                 440                 445

Cys Glu Phe Ile Ser Ala Tyr Gly Ala Arg Gly Ser Ala Asn Ile Lys
    450                 455                 460

Val Thr Phe Ile Ser Val Ala Asn Leu Thr Ile Thr Pro Asp Pro Ile
465                 470                 475                 480

Ser Val Ser Glu Gly Gln Asn Phe Ser Ile Lys Cys Ile Ser Asp Val
                485                 490                 495

Ser Asn Tyr Asp Glu Val Tyr Trp Asn Thr Ser Ala Gly Ile Lys Ile
            500                 505                 510

Tyr Gln Arg Phe Tyr Thr Thr Arg Arg Tyr Leu Asp Gly Ala Glu Ser
        515                 520                 525

Val Leu Thr Val Lys Thr Ser Thr Arg Glu Trp Asn Gly Thr Tyr His
    530                 535                 540

Cys Ile Phe Arg Tyr Lys Asn Ser Tyr Ser Ile Ala Thr Lys Asp Val
545                 550                 555                 560

Ile Val His Pro Leu Pro Leu Lys Leu Asn Ile Met Val Asp Pro Leu
```

565 570 575

Glu Ala Thr Val Ser Cys Ser Gly Ser His His Ile Lys Cys Cys Ile
580 585 590

Glu Glu Asp Gly Asp Tyr Lys Val Thr Phe His Met Gly Ser Ser Ser
595 600 605

Leu Pro Ala Ala Lys Glu Val Asn Lys Lys Gln Val Cys Tyr Lys His
610 615 620

Asn Phe Asn Ala Ser Ser Val Ser Trp Cys Ser Lys Thr Val Asp Val
625 630 635 640

Cys Cys His Phe Thr Asn Ala Ala Asn Asn Ser Val Trp Ser Pro Ser
645 650 655

Met Lys Leu Asn Leu Val Pro Gly Glu Asn Ile Thr Cys Gln Asp Pro
660 665 670

Val Ile Gly Val Gly Glu Pro Gly Lys Val Ile Gln Lys Leu Cys Arg
675 680 685

Phe Ser Asn Val Pro Ser Ser Pro Glu Ser Pro Ile Gly Gly Thr Ile
690 695 700

Thr Tyr Lys Cys Val Gly Ser Gln Trp Glu Glu Lys Arg Asn Asp Cys
705 710 715 720

Ile Ser Ala Pro Ile Asn Ser Leu Leu Gln Met Ala Lys Ala Leu Ile
725 730 735

Lys Ser Pro Ser Gln Asp Glu Met Leu Pro Thr Tyr Leu Lys Asp Leu
740 745 750

Ser Ile Ser Ile Asp Lys Ala Glu His Glu Ile Ser Ser Ser Pro Gly
755 760 765

Ser Leu Gly Ala Ile Ile Asn Ile Leu Asp Leu Leu Ser Thr Val Pro
770 775 780

Thr Gln Val Asn Ser Glu Met Met Thr His Val Leu Ser Thr Val Asn
785 790 795 800

Val Ile Leu Gly Lys Pro Val Leu Asn Thr Trp Lys Val Leu Gln Gln
805 810 815

Gln Trp Thr Asn Gln Ser Ser Gln Leu Leu His Ser Val Glu Arg Phe
820 825 830

Ser Gln Ala Leu Gln Ser Gly Asp Ser Pro Pro Leu Ser Phe Ser Gln
835 840 845

Thr Asn Val Gln Met Ser Ser Thr Val Ile Lys Ser Ser His Pro Glu
850 855 860

Thr Tyr Gln Gln Arg Phe Val Phe Pro Tyr Phe Asp Leu Trp Gly Asn
865 870 875 880

Val Val Ile Asp Lys Ser Tyr Leu Glu Asn Leu Gln Ser Asp Ser Ser
885 890 895

Ile Val Thr Met Ala Phe Pro Thr Leu Gln Ala Ile Leu Ala Gln Asp
900 905 910

Ile Gln Glu Asn Asn Phe Ala Glu Ser Leu Val Met Thr Thr Thr Val
915 920 925

Ser His Asn Thr Thr Met Pro Phe Arg Ile Ser Met Thr Phe Lys Asn
930 935 940

Asn Ser Pro Ser Gly Gly Glu Thr Lys Cys Val Phe Trp Asn Phe Arg
945 950 955 960

Leu Ala Asn Asn Thr Gly Gly Trp Asp Ser Ser Gly Cys Tyr Val Glu
965 970 975

Glu Gly Asp Gly Asp Asn Val Thr Cys Ile Cys Asp His Leu Thr Ser
980 985 990

Phe Ser Ile Leu Met Ser Pro Asp Ser Pro Asp Pro Ser Ser Leu Leu
995 1000 1005

Gly Ile Leu Leu Asp Ile Ile Ser Tyr Val Gly Val Gly Phe Ser
1010 1015 1020

Ile Leu Ser Leu Ala Ala Cys Leu Val Val Glu Ala Val Val Trp
1025 1030 1035

Lys Ser Val Thr Lys Asn Arg Thr Ser Tyr Met Arg His Thr Cys
1040 1045 1050

Ile Val Asn Ile Ala Ala Ser Leu Leu Val Ala Asn Thr Trp Phe
1055 1060 1065

Ile Val Val Ala Ala Ile Gln Asp Asn Arg Tyr Ile Leu Cys Lys
1070 1075 1080

Thr Ala Cys Val Ala Ala Thr Phe Phe Ile His Phe Phe Tyr Leu
1085 1090 1095

Ser Val Phe Phe Trp Met Leu Thr Leu Gly Leu Met Leu Phe Tyr
1100 1105 1110

Arg Leu Val Phe Ile Leu His Glu Thr Ser Arg Ser Thr Gln Lys
1115 1120 1125

Ala Ile Ala Phe Cys Leu Gly Tyr Gly Cys Pro Leu Ala Ile Ser
1130 1135 1140

Val Ile Thr Leu Gly Ala Thr Gln Pro Arg Glu Val Tyr Thr Arg
1145 1150 1155

Lys Asn Val Cys Trp Leu Asn Trp Glu Asp Thr Lys Ala Leu Leu
1160 1165 1170

Ala Phe Ala Ile Pro Ala Leu Ile Ile Val Val Val Asn Ile Thr
1175 1180 1185

Ile Thr Ile Val Val Ile Thr Lys Ile Leu Arg Pro Ser Ile Gly
1190 1195 1200

Asp Lys Pro Cys Lys Gln Glu Lys Ser Ser Leu Phe Gln Ile Ser
1205 1210 1215

Lys Ser Ile Gly Val Leu Thr Pro Leu Leu Gly Leu Thr Trp Gly
1220 1225 1230

Phe Gly Leu Thr Thr Val Phe Pro Gly Thr Asn Leu Val Phe His
1235 1240 1245

Ile Ile Phe Ala Ile Leu Asn Val Phe Gln Gly Leu Phe Ile Leu
1250 1255 1260

Leu Phe Gly Cys Leu Trp Asp Leu Lys Val Gln Glu Ala Leu Leu
1265 1270 1275

Asn Lys Phe Ser Leu Ser Arg Trp Ser Ser Gln His Ser Lys Ser
1280 1285 1290

Thr Ser Leu Gly Ser Ser Thr Pro Val Phe Ser Met Ser Ser Pro
1295 1300 1305

Ile Ser Arg Arg Phe Asn Asn Leu Phe Gly Lys Thr Gly Thr Tyr
1310 1315 1320

Asn Val Ser Thr Pro Glu Ala Thr Ser Ser Ser Leu Glu Asn Ser
1325 1330 1335

Ser Ser Ala Ser Ser Leu Leu Asn
1340 1345

<210> 21
<211> 4041
<212> DNA
<213> Homo sapiens

<400> 21
atgaaatccc caaggagaac cactttgtgc ctcatgttta ttgtgattta ttcttccaaa 60

gctgcactga actggaatta cgagtctact attcatcctt tgagtcttca tgaacatgaa 120

ccagctggtg aagaggcact gaggcaaaaa cgagccgttg ccacaaaaag tcctacggct 180

gaagaataca ctgttaatat tgagatcagt tttgaaaatg catccttcct ggatcctatc 240

aaagcctact tgaacagcct cagttttcca attcatggga ataacactga ccaaattact 300

gacattttga gcataaatgt gacaacagtc tgcagacctg ctggaaatga aatctggtgc 360

tcctgcgaga caggttatgg gtggcctcgg gaaaggtgtc ttcacaatct catttgtcaa 420

gagcgtgacg tcttcctccc agggcatcat tgcagttgcc ttaaagaact gcctcccaat 480

ggaccttttt gcctgcttca ggaagatgtt accctgaaca tgagagtcag actaaatgta 540

ggctttcaag aagacctcat gaacacttcc tccgccctct ataggtccta caagaccgac 600

ttggaaacag cgttccggaa gggttacgga attttaccag gcttcaaggg cgtgactgtg 660

acagggttca agtctggaag tgtggttgtg acatatgaag tcaagactac accaccatca 720

cttgagttaa tacataaagc caatgaacaa gttgtacaga gcctcaatca gacctacaaa 780

atggactaca actcctttca agcagttact atcaatgaaa gcaatttctt tgtcacacca 840

gaaatcatct ttgaagggga cacagtcagt ctggtgtgtg aaaaggaagt tttgtcctcc 900

aatgtgtctt ggcgctatga agaacagcag ttggaaatcc agaacagcag cagattctcg 960

atttacaccg cacttttcaa caacatgact tcggtgtcca agctcaccat ccacaacatc 1020

actccaggtg atgcaggtga atatgtttgc aaactgatat tagacatttt tgaatatgag 1080

tgcaagaaga aaatagatgt tatgcccatc caaattttgg caaatgaaga aatgaaggtg 1140

atgtgcgaca acaatcctgt atctttgaac tgctgcagtc agggtaatgt taattggagc 1200

aaagtagaat ggaagcagga aggaaaaata aatattccag gaacccctga gacagacata 1260

gattctagct gcagcagata caccctcaag gctgatggaa cccagtgccc aagcgggtcg 1320

tctggaacaa cagtcatcta cacttgtgag ttcatcagtg cctatggagc cagaggcagt 1380
gcaaacataa aagtgacatt catctctgtg gccaatctaa caataacccc ggacccaatt 1440
tctgtttctg agggacaaaa cttttctata aaatgcatca gtgatgtgag taactatgat 1500
gaggtttatt ggaacacttc tgctggaatt aaaatatacc aaagatttta taccacgagg 1560
aggtatcttg atggagcaga atcagtactg acagtcaaga cctcgaccag ggagtggaat 1620
ggaacctatc actgcatatt tagatataag aattcataca gtattgcaac caaagacgtc 1680
attgttcacc cgctgcctct aaagctgaac atcatggttg atcctttgga agctactgtt 1740
tcatgcagtg gttcccatca catcaagtgc tgcatagagg aggatggaga ctacaaagtt 1800
actttccata tgggttcctc atcccttcct gctgcaaaag aagttaacaa aaaacaagtg 1860
tgctacaaac acaatttcaa tgcaagctca gtttcctggt gttcaaaaac tgttgatgtg 1920
tgttgtcact ttaccaatgc tgctaataat tcagtttgga gcccatctat gaagctgaat 1980
ctggttcctg ggaaaacat cacatgccag gatcccgtaa taggtgtcgg agagccgggg 2040
aaagtcatcc agaagctatg ccggttctca aacgttccca gcagccctga gagtcccatt 2100
ggcgggacca tcacttacaa atgtgtaggc tcccagtggg aggagaagag aaatgactgc 2160
atctctgccc caataaacag tctgctccag atggctaagg ctttgatcaa gagcccctct 2220
caggatgaga tgctccctac atacctgaag gatctttcta ttagcataga caaagcggaa 2280
catgaaatca gctcttctcc tgggagtctg ggagccatta ttaacatcct tgatctgctc 2340
tcaacagttc caacccaagt aaattcagaa atgatgacgc acgtgctctc tacggttaat 2400
gtcatccttg gcaagcccgt cttgaacacc tggaaggttt tacaacagca atggaccaat 2460
cagagttcac agctactaca ttcagtggaa agattttccc aagcattaca gtcaggagat 2520
agccctcctt tgtccttctc ccaaactaat gtgcagatga gcagcacggt aatcaagtcc 2580
agccacccag aaacctatca acagaggttt gttttcccat actttgacct ctggggcaat 2640
gtggtcattg acaagagcta cctagaaaac ttgcagtcgg attcgtctat tgtcaccatg 2700
gctttcccaa ctctccaagc catccttgct caggatatcc aggaaaataa ctttgcagag 2760
agcttagtga tgacaaccac tgtcagccac aatacgacta tgccattcag gatttcaatg 2820
acttttaaga acaatagccc ttcaggcggc gaaacgaagt gtgtcttctg gaacttcagg 2880
cttgccaaca acacaggggg gtgggacagc agtgggtgct atgttgaaga aggtgatggg 2940
gacaatgtca cctgtatctg tgaccaccta acatcattct ccatcctcat gtcccctgac 3000
tccccagatc ctagttctct cctgggaata ctcctggata ttatttctta tgttggggtg 3060
ggcttttcca tcttgagctt ggcagcctgt ctagttgtgg aagctgtggt gtggaaatcg 3120
gtgaccaaga atcggacttc ttatatgcgc cacacctgca tagtgaatat cgctgcctcc 3180

cttctggtcg ccaacacctg gttcattgtg gtcgctgcca tccaggacaa tcgctacata 3240

ctctgcaaga cagcctgtgt ggctgccacc ttcttcatcc acttcttcta cctcagcgtc 3300

ttcttctgga tgctgacact gggcctcatg ctgttctatc gcctggtttt cattctgcat 3360

gaaacaagca ggtccactca gaaagccatt gccttctgtc ttggctatgg ctgcccactt 3420

gccatctcgg tcatcacgct gggagccacc cagccccggg aagtctatac gaggaagaat 3480

gtctgttggc tcaactggga ggacaccaag gccctgctgg ctttcgccat cccagcactg 3540

atcattgtgg tggtgaacat aaccatcact attgtggtca tcaccaagat cctgaggcct 3600

tccattggag acaagccatg caagcaggag aagagcagcc tgtttcagat cagcaagagc 3660

attggggtcc tcacaccact cttgggcctc acttggggtt ttggtctcac cactgtgttc 3720

ccagggacca accttgtgtt ccatatcata tttgccatcc tcaatgtctt ccagggatta 3780

ttcattttac tctttggatg cctctgggat ctgaaggtac aggaagcttt gctgaataag 3840

ttttcattgt cgagatggtc ttcacagcac tcaaagtcaa catccctggg ttcatccaca 3900

cctgtgtttt ctatgagttc tccaatatca aggagattta acaatttgtt tggtaaaaca 3960

ggaacgtata atgtttccac cccagaagca accagctcat ccctggaaaa ctcatccagt 4020

gcttcttcgt tgctcaacta a 4041

<210> 22
<211> 265
<212> PRT
<213> Homo sapiens

<400> 22

Gly Ile Leu Leu Asp Ile Ile Ser Tyr Val Gly Val Gly Phe Ser Ile
1 5 10 15

Leu Ser Leu Ala Ala Cys Leu Val Val Glu Ala Val Val Trp Lys Ser
20 25 30

Val Thr Lys Asn Arg Thr Ser Tyr Met Arg His Thr Cys Ile Val Asn
35 40 45

Ile Ala Ala Ser Leu Leu Val Ala Asn Thr Trp Phe Ile Val Val Ala
50 55 60

Ala Ile Gln Asp Asn Arg Tyr Ile Leu Cys Lys Thr Ala Cys Val Ala
65 70 75 80

Ala Thr Phe Phe Ile His Phe Phe Tyr Leu Ser Val Phe Phe Trp Met
85 90 95

Leu Thr Leu Gly Leu Met Leu Phe Tyr Arg Leu Val Phe Ile Leu His
100 105 110

Glu Thr Ser Arg Ser Thr Gln Lys Ala Ile Ala Phe Cys Leu Gly Tyr
115 120 125

Gly Cys Pro Leu Ala Il Ser Val Ile Thr Leu Gly Ala Thr Gln Pro

130 135 140

Arg Glu Val Tyr Thr Arg Lys Asn Val Cys Trp Leu Asn Trp Glu Asp
145 150 155 160

Thr Lys Ala Leu Leu Ala Phe Ala Ile Pro Ala Leu Ile Ile Val Val
165 170 175

Val Asn Ile Thr Ile Thr Ile Val Val Ile Thr Lys Ile Leu Arg Pro
180 185 190

Ser Ile Gly Asp Lys Pro Cys Lys Gln Glu Lys Ser Ser Leu Phe Gln
195 200 205

Ile Ser Lys Ser Ile Gly Val Leu Thr Pro Leu Leu Gly Leu Thr Trp
210 215 220

Gly Phe Gly Leu Thr Thr Val Phe Pro Gly Thr Asn Leu Val Phe His
225 230 235 240

Ile Ile Phe Ala Ile Leu Asn Val Phe Gln Gly Leu Phe Ile Leu Leu
245 250 255

Phe Gly Cys Leu Trp Asp Leu Lys Val
260 265

<210> 23
<211> 21
<212> PRT
<213> Homo sapiens

<400> 23

Met Lys Ser Pro Arg Arg Thr Thr Leu Cys Leu Met Phe Ile Val Ile
1 5 10 15

Tyr Ser Ser Lys Ala
20

<210> 24
<211> 1325
<212> PRT
<213> Homo sapiens

<400> 24

Ala Leu Asn Trp Asn Tyr Glu Ser Thr Ile His Pro Leu Ser Leu His
1 5 10 15

Glu His Glu Pro Ala Gly Glu Glu Ala Leu Arg Gln Lys Arg Ala Val
20 25 30

Ala Thr Lys Ser Pro Thr Ala Glu Glu Tyr Thr Val Asn Ile Glu Ile
35 40 45

Ser Phe Glu Asn Ala Ser Phe Leu Asp Pro Ile Lys Ala Tyr Leu Asn
50 55 60

Ser Leu Ser Phe Pro Ile His Gly Asn Asn Thr Asp Gln Ile Thr Asp
65 70 75 80

Ile Leu Ser Ile Asn Val Thr Thr Val Cys Arg Pro Ala Gly Asn Glu
85 90 95

Ile Trp Cys Ser Cys Glu Thr Gly Tyr Gly Trp Pro Arg Glu Arg Cys
100 105 110

Leu His Asn Leu Ile Cys Gln Glu Arg Asp Val Phe Leu Pro Gly His
115 120 125

His Cys Ser Cys Leu Lys Glu Leu Pro Pro Asn Gly Pro Phe Cys Leu
130 135 140

Leu Gln Glu Asp Val Thr Leu Asn Met Arg Val Arg Leu Asn Val Gly
145 150 155 160

Phe Gln Glu Asp Leu Met Asn Thr Ser Ser Ala Leu Tyr Arg Ser Tyr
165 170 175

Lys Thr Asp Leu Glu Thr Ala Phe Arg Lys Gly Tyr Gly Ile Leu Pro
180 185 190

Gly Phe Lys Gly Val Thr Val Thr Gly Phe Lys Ser Gly Ser Val Val
195 200 205

Val Thr Tyr Glu Val Lys Thr Thr Pro Pro Ser Leu Glu Leu Ile His
210 215 220

Lys Ala Asn Glu Gln Val Val Gln Ser Leu Asn Gln Thr Tyr Lys Met
225 230 235 240

Asp Tyr Asn Ser Phe Gln Ala Val Thr Ile Asn Glu Ser Asn Phe Phe
245 250 255

Val Thr Pro Glu Ile Ile Phe Glu Gly Asp Thr Val Ser Leu Val Cys
260 265 270

Glu Lys Glu Val Leu Ser Ser Asn Val Ser Trp Arg Tyr Glu Glu Gln
275 280 285

Gln Leu Glu Ile Gln Asn Ser Ser Arg Phe Ser Ile Tyr Thr Ala Leu
290 295 300

Phe Asn Asn Met Thr Ser Val Ser Lys Leu Thr Ile His Asn Ile Thr
305 310 315 320

Pro Gly Asp Ala Gly Glu Tyr Val Cys Lys Leu Ile Leu Asp Ile Phe
325 330 335

Glu Tyr Glu Cys Lys Lys Lys Ile Asp Val Met Pro Ile Gln Ile Leu
340 345 350

Ala Asn Glu Glu Met Lys Val Met Cys Asp Asn Asn Pro Val Ser Leu
355 360 365

Asn Cys Cys Ser Gln Gly Asn Val Asn Trp Ser Lys Val Glu Trp Lys
370 375 380

Gln Glu Gly Lys Ile Asn Ile Pro Gly Thr Pro Glu Thr Asp Ile Asp
385 390 395 400

Ser Ser Cys Ser Arg Tyr Thr Leu Lys Ala Asp Gly Thr Gln Cys Pro
405 410 415

Ser Gly Ser Ser Gly Thr Thr Val Ile Tyr Thr Cys Glu Phe Ile Ser
420 425 430

Ala Tyr Gly Ala Arg Gly Ser Ala Asn Ile Lys Val Thr Phe Ile Ser
435 440 445

Val Ala Asn Leu Thr Ile Thr Pro Asp Pro Ile Ser Val Ser Glu Gly
450 455 460

Gln Asn Phe Ser Ile Lys Cys Ile Ser Asp Val Ser Asn Tyr Asp Glu
465 470 475 480

Val Tyr Trp Asn Thr Ser Ala Gly Ile Lys Ile Tyr Gln Arg Phe Tyr
485 490 495

Thr Thr Arg Arg Tyr Leu Asp Gly Ala Glu Ser Val Leu Thr Val Lys
500 505 510

Thr Ser Thr Arg Glu Trp Asn Gly Thr Tyr His Cys Ile Phe Arg Tyr
515 520 525

Lys Asn Ser Tyr Ser Ile Ala Thr Lys Asp Val Ile Val His Pro Leu
530 535 540

Pro Leu Lys Leu Asn Ile Met Val Asp Pro Leu Glu Ala Thr Val Ser
545 550 555 560

Cys Ser Gly Ser His His Ile Lys Cys Cys Ile Glu Glu Asp Gly Asp
565 570 575

Tyr Lys Val Thr Phe His Met Gly Ser Ser Ser Leu Pro Ala Ala Lys
580 585 590

Glu Val Asn Lys Lys Gln Val Cys Tyr Lys His Asn Phe Asn Ala Ser
595 600 605

Ser Val Ser Trp Cys Ser Lys Thr Val Asp Val Cys Cys His Phe Thr
610 615 620

Asn Ala Ala Asn Asn Ser Val Trp Ser Pro Ser Met Lys Leu Asn Leu
625 630 635 640

Val Pro Gly Glu Asn Ile Thr Cys Gln Asp Pro Val Ile Gly Val Gly
645 650 655

Glu Pro Gly Lys Val Ile Gln Lys Leu Cys Arg Phe Ser Asn Val Pro
660 665 670

Ser Ser Pro Glu Ser Pro Ile Gly Gly Thr Ile Thr Tyr Lys Cys Val
675 680 685

Gly Ser Gln Trp Glu Glu Lys Arg Asn Asp Cys Ile Ser Ala Pro Ile
690 695 700

Asn Ser Leu Leu Gln Met Ala Lys Ala Leu Ile Lys Ser Pro Ser Gln
705 710 715 720

Asp Glu Met Leu Pro Thr Tyr Leu Lys Asp Leu Ser Ile Ser Ile Asp
725 730 735

Lys Ala Glu His Glu Ile Ser Ser Ser Pro Gly Ser Leu Gly Ala Ile
740 745 750

Ile Asn Ile Leu Asp Leu Leu Ser Thr Val Pro Thr Gln Val Asn Ser
755 760 765

Glu Met Met Thr His Val Leu Ser Thr Val Asn Val Ile Leu Gly Lys
770 775 780

Pro Val Leu Asn Thr Trp Lys Val Leu Gln Gln Gln Trp Thr Asn Gln
785 790 795 800

Ser Ser Gln Leu Leu His Ser Val Glu Arg Phe Ser Gln Ala Leu Gln
805 810 815

Ser Gly Asp Ser Pro Pro Leu Ser Phe Ser Gln Thr Asn Val Gln Met
820 825 830

Ser Ser Thr Val Ile Lys Ser Ser His Pro Glu Thr Tyr Gln Gln Arg
835 840 845

Phe Val Phe Pro Tyr Phe Asp Leu Trp Gly Asn Val Val Ile Asp Lys
850 855 860

Ser Tyr Leu Glu Asn Leu Gln Ser Asp Ser Ser Ile Val Thr Met Ala
865 870 875 880

Phe Pro Thr Leu Gln Ala Ile Leu Ala Gln Asp Ile Gln Glu Asn Asn
885 890 895

Phe Ala Glu Ser Leu Val Met Thr Thr Thr Val Ser His Asn Thr Thr
900 905 910

Met Pro Phe Arg Ile Ser Met Thr Phe Lys Asn Asn Ser Pro Ser Gly
915 920 925

Gly Glu Thr Lys Cys Val Phe Trp Asn Phe Arg Leu Ala Asn Asn Thr
930 935 940

Gly Gly Trp Asp Ser Ser Gly Cys Tyr Val Glu Glu Gly Asp Gly Asp
945 950 955 960

Asn Val Thr Cys Ile Cys Asp His Leu Thr Ser Phe Ser Ile Leu Met
965 970 975

Ser Pro Asp Ser Pro Asp Pro Ser Ser Leu Leu Gly Ile Leu Leu Asp
980 985 990

Ile Ile Ser Tyr Val Gly Val Gly Phe Ser Ile Leu Ser Leu Ala Ala
995 1000 1005

Cys Leu Val Val Glu Ala Val Val Trp Lys Ser Val Thr Lys Asn
1010 1015 1020

Arg Thr Ser Tyr Met Arg His Thr Cys Ile Val Asn Ile Ala Ala
1025 1030 1035

Ser Leu Leu Val Ala Asn Thr Trp Phe Ile Val Val Ala Ala Ile
1040 1045 1050

Gln Asp Asn Arg Tyr Ile Leu Cys Lys Thr Ala Cys Val Ala Ala
1055 1060 1065

Thr Phe Phe Ile His Phe Phe Tyr Leu Ser Val Phe Phe Trp Met
1070 1075 1080

Leu Thr Leu Gly Leu Met Leu Phe Tyr Arg Leu Val Phe Ile Leu
1085 1090 1095

His Glu Thr Ser Arg Ser Thr Gln Lys Ala Ile Ala Phe Cys Leu
1100 1105 1110

Gly Tyr Gly Cys Pro Leu Ala Ile Ser Val Ile Thr Leu Gly Ala
1115 1120 1125

Thr Gln Pro Arg Glu Val Tyr Thr Arg Lys Asn Val Cys Trp Leu
1130 1135 1140

Asn Trp Glu Asp Thr Lys Ala Leu Leu Ala Phe Ala Ile Pro Ala
1145 1150 1155

Leu Ile Ile Val Val Val Asn Ile Thr Ile Thr Ile Val Val Ile
1160 1165 1170

Thr Lys Ile Leu Arg Pro Ser Ile Gly Asp Lys Pro Cys Lys Gln
1175 1180 1185

Glu Lys Ser Ser Leu Phe Gln Ile Ser Lys Ser Ile Gly Val Leu
1190 1195 1200

Thr Pro Leu Leu Gly Leu Thr Trp Gly Phe Gly Leu Thr Thr Val
1205 1210 1215

Phe Pro Gly Thr Asn Leu Val Phe His Ile Ile Phe Ala Ile Leu
1220 1225 1230

Asn Val Phe Gln Gly Leu Phe Ile Leu Leu Phe Gly Cys Leu Trp
1235 1240 1245

Asp Leu Lys Val Gln Glu Ala Leu Leu Asn Lys Phe Ser Leu Ser
1250 1255 1260

Arg Trp Ser Ser Gln His Ser Lys Ser Thr Ser Leu Gly Ser Ser
1265 1270 1275

Thr Pro Val Phe Ser Met Ser Ser Pro Ile Ser Arg Arg Phe Asn
1280 1285 1290

Asn Leu Phe Gly Lys Thr Gly Thr Tyr Asn Val Ser Thr Pro Glu
1295 1300 1305

Ala Thr Ser Ser Ser Leu Glu Asn Ser Ser Ser Ala Ser Ser Leu
1310 1315 1320

Leu Asn
1325

<210> 25
<211> 1371
<212> PRT
<213> Homo sapiens

<400> 25

Ala Ser Asp Gln Ser Gly Ser Gln Pro Gly Asp His Ser Ala Gly Gln
1 5 10 15

Ala Asn Gln Leu Lys Leu Glu Asp Met Lys Ser Pro Arg Arg Thr Thr
20 25 30

Leu Cys Leu Met Phe Ile Val Il Tyr Ser Ser Lys Ala Ala Leu Asn

35 40 45

Trp Asn Tyr Glu Ser Thr Ile His Pro Leu Ser Leu His Glu His Glu
50 55 60

Pro Ala Gly Glu Glu Ala Leu Arg Gln Lys Arg Ala Val Ala Thr Lys
65 70 75 80

Ser Pro Thr Ala Glu Glu Tyr Thr Val Asn Ile Glu Ile Ser Phe Glu
85 90 95

Asn Ala Ser Phe Leu Asp Pro Ile Lys Ala Tyr Leu Asn Ser Leu Ser
100 105 110

Phe Pro Ile His Gly Asn Asn Thr Asp Gln Ile Thr Asp Ile Leu Ser
115 120 125

Ile Asn Val Thr Thr Val Cys Arg Pro Ala Gly Asn Glu Ile Trp Cys
130 135 140

Ser Cys Glu Thr Gly Tyr Gly Trp Pro Arg Glu Arg Cys Leu His Asn
145 150 155 160

Leu Ile Cys Gln Glu Arg Asp Val Phe Leu Pro Gly His His Cys Ser
165 170 175

Cys Leu Lys Glu Leu Pro Pro Asn Gly Pro Phe Cys Leu Leu Gln Glu
180 185 190

Asp Val Thr Leu Asn Met Arg Val Arg Leu Asn Val Gly Phe Gln Glu
195 200 205

Asp Leu Met Asn Thr Ser Ser Ala Leu Tyr Arg Ser Tyr Lys Thr Asp
210 215 220

Leu Glu Thr Ala Phe Arg Lys Gly Tyr Gly Ile Leu Pro Gly Phe Lys
225 230 235 240

Gly Val Thr Val Thr Gly Phe Lys Ser Gly Ser Val Val Val Thr Tyr
245 250 255

Glu Val Lys Thr Thr Pro Pro Ser Leu Glu Leu Ile His Lys Ala Asn
260 265 270

Glu Gln Val Val Gln Ser Leu Asn Gln Thr Tyr Lys Met Asp Tyr Asn
275 280 285

Ser Phe Gln Ala Val Thr Ile Asn Glu Ser Asn Phe Phe Val Thr Pro
290 295 300

Glu Ile Ile Phe Glu Gly Asp Thr Val Ser Leu Val Cys Glu Lys Glu
305 310 315 320

Val Leu Ser Ser Asn Val Ser Trp Arg Tyr Glu Glu Gln Gln Leu Glu
325 330 335

Ile Gln Asn Ser Ser Arg Phe Ser Ile Tyr Thr Ala Leu Phe Asn Asn
340 345 350

Met Thr Ser Val Ser Lys Leu Thr Ile His Asn Ile Thr Pro Gly Asp
355 360 365

Ala Gly Glu Tyr Val Cys Lys Leu Ile Leu Asp Ile Phe Glu Tyr Glu

370 375 380

Cys Lys Lys Lys Ile Asp Val Met Pro Ile Gln Ile Leu Ala Asn Glu
385 390 395 400

Glu Met Lys Val Met Cys Asp Asn Asn Pro Val Ser Leu Asn Cys Cys
405 410 415

Ser Gln Gly Asn Val Asn Trp Ser Lys Val Glu Trp Lys Gln Glu Gly
420 425 430

Lys Ile Asn Ile Pro Gly Thr Pro Glu Thr Asp Ile Asp Ser Ser Cys
435 440 445

Ser Arg Tyr Thr Leu Lys Ala Asp Gly Thr Gln Cys Pro Ser Gly Ser
450 455 460

Ser Gly Thr Thr Val Ile Tyr Thr Cys Glu Phe Ile Ser Ala Tyr Gly
465 470 475 480

Ala Arg Gly Ser Ala Asn Ile Lys Val Thr Phe Ile Ser Val Ala Asn
485 490 495

Leu Thr Ile Thr Pro Asp Pro Ile Ser Val Ser Glu Gly Gln Asn Phe
500 505 510

Ser Ile Lys Cys Ile Ser Asp Val Ser Asn Tyr Asp Glu Val Tyr Trp
515 520 525

Asn Thr Ser Ala Gly Ile Lys Ile Tyr Gln Arg Phe Tyr Thr Thr Arg
530 535 540

Arg Tyr Leu Asp Gly Ala Glu Ser Val Leu Thr Val Lys Thr Ser Thr
545 550 555 560

Arg Glu Trp Asn Gly Thr Tyr His Cys Ile Phe Arg Tyr Lys Asn Ser
565 570 575

Tyr Ser Ile Ala Thr Lys Asp Val Ile Val His Pro Leu Pro Leu Lys
580 585 590

Leu Asn Ile Met Val Asp Pro Leu Glu Ala Thr Val Ser Cys Ser Gly
595 600 605

Ser His His Ile Lys Cys Cys Ile Glu Glu Asp Gly Asp Tyr Lys Val
610 615 620

Thr Phe His Met Gly Ser Ser Ser Leu Pro Ala Ala Lys Glu Val Asn
625 630 635 640

Lys Lys Gln Val Cys Tyr Lys His Asn Phe Asn Ala Ser Ser Val Ser
645 650 655

Trp Cys Ser Lys Thr Val Asp Val Cys Cys His Phe Thr Asn Ala Ala
660 665 670

Asn Asn Ser Val Trp Ser Pro Ser Met Lys Leu Asn Leu Val Pro Gly
675 680 685

Glu Asn Ile Thr Cys Gln Asp Pro Val Ile Gly Val Gly Glu Pro Gly
690 695 700

Lys Val Ile Gln Lys Leu Cys Arg Phe Ser Asn Val Pro Ser Ser Pro

705 710 715 720

Glu Glu Ser Pro Leu Gly Gly Thr Ile Thr Tyr Lys Cys Val Gly Ser
725 730 735

Gln Trp Gly Glu Lys Arg Asn Asp Cys Ile Ser Ala Pro Ile Asn Ser
740 745 750

Leu Leu Gln Met Ala Lys Ala Leu Ile Lys Ser Pro Ser Gln Asp Glu
755 760 765

Met Leu Pro Thr Tyr Leu Lys Asp Leu Ser Ile Ser Ile Asp Lys Ala
770 775 780

Glu His Glu Ile Ser Ser Ser Pro Gly Ser Leu Gly Ala Ile Ile Asn
785 790 795 800

Ile Leu Asp Leu Leu Ser Thr Val Pro Thr Gln Val Asn Ser Glu Met
805 810 815

Met Thr His Val Leu Ser Thr Val Asn Val Ile Leu Gly Lys Pro Val
820 825 830

Leu Asn Thr Trp Lys Val Leu Gln Gln Gln Trp Thr Asn Gln Ser Ser
835 840 845

Gln Leu Leu His Ser Val Glu Arg Phe Ser Gln Ala Leu Gln Ser Gly
850 855 860

Asp Ser Pro Pro Leu Ser Phe Ser Gln Thr Asn Val Gln Met Ser Ser
865 870 875 880

Thr Val Ile Lys Ser Ser His Pro Glu Thr Tyr Gln Gln Arg Phe Val
885 890 895

Phe Pro Tyr Phe Asp Leu Trp Gly Asn Val Val Ile Asp Lys Ser Tyr
900 905 910

Leu Glu Asn Leu Gln Ser Asp Ser Ser Ile Val Thr Met Ala Phe Pro
915 920 925

Thr Leu Gln Ala Ile Leu Ala Gln Asp Ile Gln Glu Asn Asn Phe Ala
930 935 940

Glu Ser Leu Val Met Thr Thr Thr Val Ser His Asn Thr Thr Met Pro
945 950 955 960

Phe Arg Ile Ser Met Thr Phe Lys Asn Asn Ser Pro Ser Gly Gly Glu
965 970 975

Thr Lys Cys Val Phe Trp Asn Phe Arg Leu Ala Asn Asn Thr Gly Gly
980 985 990

Trp Asp Ser Ser Gly Cys Tyr Val Glu Glu Gly Asp Gly Asp Asn Val
995 1000 1005

Thr Cys Ile Cys Asp His Leu Thr Ser Phe Ser Ile Leu Met Ser
1010 1015 1020

Pro Asp Ser Pro Asp Pro Ser Ser Leu Leu Gly Ile Leu Leu Asp
1025 1030 1035

Ile Ile Ser Tyr Val Gly Val Gly Phe Ser Ile Leu Ser Leu Ala

1040 1045 1050

Ala Cys Leu Val Val Glu Ala Val Val Trp Lys Ser Val Thr Lys
1055 1060 1065

Asn Arg Thr Ser Tyr Met Arg His Thr Cys Ile Val Asn Ile Ala
1070 1075 1080

Ala Ser Leu Leu Val Ala Asn Thr Trp Phe Ile Gly Val Ala Ala
1085 1090 1095

Ile Gln Asp Asn Arg Tyr Ile Leu Cys Lys Thr Ala Cys Val Ala
1100 1105 1110

Ala Thr Phe Phe Ile His Phe Phe Tyr Leu Ser Val Phe Phe Trp
1115 1120 1125

Met Leu Thr Leu Gly Leu Met Leu Phe Tyr Arg Leu Val Phe Ile
1130 1135 1140

Leu His Glu Thr Ser Arg Ser Thr Gln Lys Ala Ile Ala Phe Cys
1145 1150 1155

Leu Gly Tyr Gly Cys Pro Leu Ala Ile Ser Val Ile Thr Leu Gly
1160 1165 1170

Ala Thr Gln Pro Arg Glu Val Tyr Thr Arg Lys Asn Val Cys Trp
1175 1180 1185

Leu Asn Trp Glu Asp Thr Lys Ala Leu Leu Ala Phe Ala Ile Pro
1190 1195 1200

Ala Leu Ile Ile Val Val Val Asn Ile Thr Ile Thr Ile Val Val
1205 1210 1215

Ile Thr Lys Ile Leu Arg Pro Ser Ile Gly Asp Lys Pro Cys Lys
1220 1225 1230

Gln Glu Lys Ser Ser Leu Phe Gln Ile Ser Lys Ser Ile Gly Val
1235 1240 1245

Leu Thr Pro Leu Leu Gly Leu Thr Trp Gly Phe Gly Leu Thr Thr
1250 1255 1260

Val Phe Pro Gly Thr Asn Leu Val Phe His Ile Ile Phe Ala Ile
1265 1270 1275

Leu Asn Val Phe Gln Gly Leu Phe Ile Leu Leu Phe Gly Cys Leu
1280 1285 1290

Trp Asp Leu Lys Val Gln Glu Ala Leu Leu Asn Lys Phe Ser Leu
1295 1300 1305

Ser Arg Trp Ser Ser Gln His Ser Lys Ser Thr Ser Leu Gly Ser
1310 1315 1320

Ser Thr Pro Val Phe Ser Met Ser Ser Pro Ile Ser Arg Arg Phe
1325 1330 1335

Asn Asn Leu Phe Gly Lys Thr Gly Thr Tyr Asn Val Ser Thr Pro
1340 1345 1350

Glu Ala Thr Ser Ser Ser Leu Glu Asn Ser Ser Ser Ala Ser Ser

1355 1360 1365

Leu Leu Asn
1370

<210> 26
<211> 386
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(386)
<223> n = A, T, G, or C

<400> 26
gagagtaacc tgtctggcct ggtgcctgct gccgggctgg tgcctgcgct gccacctgct 60

gtgaccctgg ggctgacagc tgcctacacc accctgtatg ccctgctctt cttctccgtc 120

tatgcccagc tctggctggt gcttctgtat gggcacaagc gtctcagcta tcaaacgggg 180

ttcctggccc tctgtctgct ctgggccgcc ttgcgtacca ccctcttctc cttctacttc 240

cgagatactc cccgcgccaa ccgcctgggg cccttgccct tctggcttct ctactgctgc 300

cccgtctgcc tgcagttctt caccttgacg cttatgaacc tctactttgc ccaggtggtg 360

ttcaaggcca aagtgaagcg tcngcc 386

<210> 27
<211> 1372
<212> DNA
<213> Homo sapiens

<400> 27
gggccgcctg gacacacagg tcccctggca caggtaacaa cagaggaccc tcagggagag 60

aaggctccca gcttcctcta caccctgaga accgctgccg agagacccaa aagacacctg 120

ggcatcagtg ctgcagccct accttggcct ccaggtagat gctagtggag ggcgcccgcc 180

tggcgacaaa ggaagaggca ggcagcaaga gacagcgcgc agatgacgaa cagggagtcg 240

ctcaccagga cgcggacaag cagcagggcc cagggctgtg ccgccgcgcc gcatgggaga 300

gcacagcaca cagcacgttc accagcagaa agagcagcga ggccccccaca aaggcccctc 360

ggacagcgag caagcctcgg ctcatcttcg ggcccgacgc ttcactttgg ccttgaacac 420

cacctggggc aaagtagagg ttcataaggc gtcaaggtga agaactgcag gcagacgggg 480

cagcagtaga gaagccagaa gggcaaaggg ccccagggtt gtttggcgcg gggagtatct 540

tcggaaatta gaaggagaaa aaggttgtta cgcaaggggg ccccagaagc agacagaggg 600

ccaggaacac ccgtctgata agctgagacg cttgtgcccc atacagaagc accagccaga 660

gctgggcata gacggagaag aagagcaggg catacagggt ggtgtaggca gctgtcagcc 720

ccagggtcac agcaggtggc agcgcaggca ccaagcccgg caagcaggca accaggccca 780

```
gacaggttac tctccatgtc aagggaggga ggcctgaatc ccggagacag acaaggaggg      840

actgtcggcg ggaccaggga aataaatacg gaggggtggg aattgcaggg tgcagagagg      900

gagatggggc gctcaggagg aaagaagttg tggctgacag gcccccctcc tcaacagaga      960

gacccttgcc cttggggaac acaggcagcc tgggccccag gcccctcact caggggcatg     1020

gaagggcagt ctcttttctc caggacagac cggatagggg ttgggggtgt ctcctccctg     1080

ggcctacccc agagccaagg atggggggat gcagaggccc aaggagagga ctccctggcc     1140

tcggtcgttc atgtcgtcct gggtggaggg ggaggctctc tctcctccaa taatgcggat     1200

ggggagggcc tctgcctcct gtcctcgtaa ttctcgaaag ctgcccccgc cccaccctcc     1260

cccaccgcgt ccacgctcca gcctcaaggg gagggaggag gcggacgcca ggtcctcagg     1320

ccgctgatgg gagagccggc tcaggtcgtg ggacccgggg gctccgacga aa             1372


<210>  28
<211>  3996
<212>  DNA
<213>  Homo sapiens

<220>
<221>  CDS
<222>  (52)..(3996)

<400>  28
gccgacagtc cctccttgtc tgtctccggg attcaggcct ccctccctga c atg gag        57
                                                         Met Glu
                                                         1

agt aac ctg tct ggc ctg gtg cct gct gcc ggg ctg gtg cct gcg ctg        105
Ser Asn Leu Ser Gly Leu Val Pro Ala Ala Gly Leu Val Pro Ala Leu
        5                   10                  15

cca cct gct gtg acc ctg ggg ctg aca gct gcc tac acc acc ctg tat        153
Pro Pro Ala Val Thr Leu Gly Leu Thr Ala Ala Tyr Thr Thr Leu Tyr
    20                  25                  30

gcc ctg ctc ttc ttc tcc gtc tat gcc cag ctc tgg ctg gtg ctt ctg        201
Ala Leu Leu Phe Phe Ser Val Tyr Ala Gln Leu Trp Leu Val Leu Leu
35                  40                  45                  50

tat ggg cac aag cgt ctc agc tat cag acg gtg ttc ctg gcc ctc tgt        249
Tyr Gly His Lys Arg Leu Ser Tyr Gln Thr Val Phe Leu Ala Leu Cys
                55                  60                  65

ctg ctc tgg gcc gcc ttg cgt acc acc ctc ttc tcc ttc tac ttc cga        297
Leu Leu Trp Ala Ala Leu Arg Thr Thr Leu Phe Ser Phe Tyr Phe Arg
            70                  75                  80

gat act ccc cgc gcc aac cgc ctg ggg ccc ttg ccc ttc tgg ctt ctc        345
Asp Thr Pro Arg Ala Asn Arg Leu Gly Pro Leu Pro Phe Trp Leu Leu
        85                  90                  95

tac tgc tgc ccc gtc tgc ctg cag ttc ttc acc ttg acg ctt atg aac        393
Tyr Cys Cys Pro Val Cys Leu Gln Phe Phe Thr Leu Thr Leu Met Asn
    100                 105                 110
```

```
ctc tac ttt gcc cag gtg gtg ttc aag gcc aag gtg aag cgt cgg ccg       441
Leu Tyr Phe Ala Gln Val Val Phe Lys Ala Lys Val Lys Arg Arg Pro
115                 120                 125                 130

gag atg agc cga ggc ttg ctc gct gtc cga ggg gcc ttt gtg ggg gcc       489
Glu Met Ser Arg Gly Leu Leu Ala Val Arg Gly Ala Phe Val Gly Ala
                135                 140                 145

tcg ctg ctc ttt ctg ctg gtg aac gtg ctg tgt gct gtg ctc tcc cat       537
Ser Leu Leu Phe Leu Leu Val Asn Val Leu Cys Ala Val Leu Ser His
            150                 155                 160

cgg cgc cgg gca cag ccc tgg gcc ctg ctg ctt gtc cgc gtc ctg gtg       585
Arg Arg Arg Ala Gln Pro Trp Ala Leu Leu Leu Val Arg Val Leu Val
        165                 170                 175

agc gac tcc ctg ttc gtc atc tgc gcg ctg tct ctt gct gcc tgc ctc       633
Ser Asp Ser Leu Phe Val Ile Cys Ala Leu Ser Leu Ala Ala Cys Leu
    180                 185                 190

tgc ctc gtc gcc agg cgg gcg ccc tcc act agc atc tac ctg gag gcc       681
Cys Leu Val Ala Arg Arg Ala Pro Ser Thr Ser Ile Tyr Leu Glu Ala
195                 200                 205                 210

aag ggg acc agt gtg tgc cag gcg gcc gcg atg ggt ggc gcc atg gtc       729
Lys Gly Thr Ser Val Cys Gln Ala Ala Ala Met Gly Gly Ala Met Val
                215                 220                 225

ctg ctc tat gcc agc cgg gcc tgc tac aac ctg aca gca ctg gcc ttg       777
Leu Leu Tyr Ala Ser Arg Ala Cys Tyr Asn Leu Thr Ala Leu Ala Leu
            230                 235                 240

gcc ccc cag agc cgg ctg gac acc ttc gat tac gac tgg tac aat gtg       825
Ala Pro Gln Ser Arg Leu Asp Thr Phe Asp Tyr Asp Trp Tyr Asn Val
        245                 250                 255

tct gac cag gcg gac ctg gtg aat gac ctg ggg aac aaa ggc tac ctg       873
Ser Asp Gln Ala Asp Leu Val Asn Asp Leu Gly Asn Lys Gly Tyr Leu
    260                 265                 270

gta ttt ggc ctc atc ctc ttc gtg tgg gag cta ctg ccc acc acc ctg       921
Val Phe Gly Leu Ile Leu Phe Val Trp Glu Leu Leu Pro Thr Thr Leu
275                 280                 285                 290

ctg gtg ggc ttc ttc cgg gtg cac cgg ccc cca cag gac ctg agc acc       969
Leu Val Gly Phe Phe Arg Val His Arg Pro Pro Gln Asp Leu Ser Thr
                295                 300                 305

agc cac atc ctc aat ggg cag gtc ttt gcc tct cgg tcc tac ttc ttt      1017
Ser His Ile Leu Asn Gly Gln Val Phe Ala Ser Arg Ser Tyr Phe Phe
            310                 315                 320

gac cgg gct ggg cac tgt gaa gat gag ggc tgc tcc tgg gag cac agc      1065
Asp Arg Ala Gly His Cys Glu Asp Glu Gly Cys Ser Trp Glu His Ser
        325                 330                 335

cgg ggt gag agc acc agc tct tgt gac tgt ggc cct ggt cac tgt cct      1113
Arg Gly Glu Ser Thr Ser Ser Cys Asp Cys Gly Pro Gly His Cys Pro
    340                 345                 350

gag act gac cct gtt tct ctg ctg cag tat gtc ggg cag tct agg ctc      1161
Glu Thr Asp Pro Val Ser Leu Leu Gln Tyr Val Gly Gln Ser Arg Leu
```

355 360 365 370

tgg gag ctg aat acc cag gcc cca gtc ccc ctc acc cta ggc ccc tgt 1209
Trp Glu Leu Asn Thr Gln Ala Pro Val Pro Leu Thr Leu Gly Pro Cys
375 380 385

gcc aag ttt gtc tgc cgc ttc ttg ccc agg atc ctg ggg gtc gtg gct 1257
Ala Lys Phe Val Cys Arg Phe Leu Pro Arg Ile Leu Gly Val Val Ala
390 395 400

acc ccc tcc tct ggc cgg ctc ctt gct gct cct gtc ata gac tca ggg 1305
Thr Pro Ser Ser Gly Arg Leu Leu Ala Ala Pro Val Ile Asp Ser Gly
405 410 415

gca ggg acg ccc cag ggg cga ctg gcc ggg cgg ggt gcc cac ctc tcg 1353
Ala Gly Thr Pro Gln Gly Arg Leu Ala Gly Arg Gly Ala His Leu Ser
420 425 430

cgc gtg ggc gcc tcc ggg agt ggt gtg gcc gcc ggt ccc gcc gcc cgc 1401
Arg Val Gly Ala Ser Gly Ser Gly Val Ala Ala Gly Pro Ala Ala Arg
435 440 445 450

cac gct ccg agg cgt cgc tgt gcg gac gcg ggg gag gcg gtg gga gcg 1449
His Ala Pro Arg Arg Arg Cys Ala Asp Ala Gly Glu Ala Val Gly Ala
455 460 465

agc tgc ggg cgc tgc gcg gtg gcc ctg ctg tct ggc gtg tgc acg cta 1497
Ser Cys Gly Arg Cys Ala Val Ala Leu Leu Ser Gly Val Cys Thr Leu
470 475 480

gtg tcc aca cac gtg tgc gtg ggc tct ggg tgc cct ggc gcg gcc ggc 1545
Val Ser Thr His Val Cys Val Gly Ser Gly Cys Pro Gly Ala Ala Gly
485 490 495

acg ccc atg ggg gcc ggg gat gcc ggg gcg tct gcg gag agt gca gtg 1593
Thr Pro Met Gly Ala Gly Asp Ala Gly Ala Ser Ala Glu Ser Ala Val
500 505 510

acg aca gct ccc cag gag ccc ccc gcc cgg ccc ctc cag gcg ggc agt 1641
Thr Thr Ala Pro Gln Glu Pro Pro Ala Arg Pro Leu Gln Ala Gly Ser
515 520 525 530

gga gct ggc ccg gcg cct ggg cgc gcc atg cgc agc acc acg ctc ctg 1689
Gly Ala Gly Pro Ala Pro Gly Arg Ala Met Arg Ser Thr Thr Leu Leu
535 540 545

gcc ctg ctg gcg ctg gtc ttg ctt tac ttg gtg tct ggt gcc ctg gtg 1737
Ala Leu Leu Ala Leu Val Leu Leu Tyr Leu Val Ser Gly Ala Leu Val
550 555 560

ttc cgg gcc ctg gag cag ccc cac gag cag cag gcc cag agg gag ctg 1785
Phe Arg Ala Leu Glu Gln Pro His Glu Gln Gln Ala Gln Arg Glu Leu
565 570 575

ggg gag gtc cga gag aag ttc ctg agg gcc cat ccg tgt gtg agc gac 1833
Gly Glu Val Arg Glu Lys Phe Leu Arg Ala His Pro Cys Val Ser Asp
580 585 590

cag gag ctg ggc ctc ctc atc aag gag gtg gct gat gcc ctg gga ggg 1881
Gln Glu Leu Gly Leu Leu Ile Lys Glu Val Ala Asp Ala Leu Gly Gly
595 600 605 610

ggt gcg gac cca gaa acc aac tcg acc agc aac agc agc cac tca gcc 1929

```
Gly Ala Asp Pro Glu Thr Asn Ser Thr Ser Asn Ser Ser His Ser Ala
                615                 620                 625

tgg gac ctg ggc agc gcc ttc ttt ttc tca ggg acc atc atc acc acc      1977
Trp Asp Leu Gly Ser Ala Phe Phe Phe Ser Gly Thr Ile Ile Thr Thr
            630                 635                 640

atc ggc tat ggc aat gtg gcc ctg cgc aca gat gcc ggg cgc ctc ttc      2025
Ile Gly Tyr Gly Asn Val Ala Leu Arg Thr Asp Ala Gly Arg Leu Phe
        645                 650                 655

tgc atc ttt tat gcg ctg gtg ggg att ccg ctg ttt ggg atc cta ctg      2073
Cys Ile Phe Tyr Ala Leu Val Gly Ile Pro Leu Phe Gly Ile Leu Leu
    660                 665                 670

gca ggg gtc ggg gac cgg ctg ggc tcc tcc ctg cgc cat ggc atc ggt      2121
Ala Gly Val Gly Asp Arg Leu Gly Ser Ser Leu Arg His Gly Ile Gly
675                 680                 685                 690

cac att gaa gcc atc ttc ttg aag tgg cac gtg cca ccg gag cta gta      2169
His Ile Glu Ala Ile Phe Leu Lys Trp His Val Pro Pro Glu Leu Val
                695                 700                 705

aga gtg ctg tcg gcg atg ctt ttc ctg ctg atc ggc tgc ctg ctc ttt      2217
Arg Val Leu Ser Ala Met Leu Phe Leu Leu Ile Gly Cys Leu Leu Phe
            710                 715                 720

gtc ctc acg ccc acg ttc gtg ttc tgc tat atg gag gac tgg agc aag      2265
Val Leu Thr Pro Thr Phe Val Phe Cys Tyr Met Glu Asp Trp Ser Lys
        725                 730                 735

ctg gag gcc atc tac ttt gtc ata gtg acg ctt acc acc gtg ggc ttt      2313
Leu Glu Ala Ile Tyr Phe Val Ile Val Thr Leu Thr Thr Val Gly Phe
    740                 745                 750

ggc gac tat gtg gcc ggc gcg gac ccc agg cag gac tcc ccg gcc tat      2361
Gly Asp Tyr Val Ala Gly Ala Asp Pro Arg Gln Asp Ser Pro Ala Tyr
755                 760                 765                 770

cag ccg ctg gtg tgg ttc tgg atc ctg ctc ggc ctg gct tac ttc gcc      2409
Gln Pro Leu Val Trp Phe Trp Ile Leu Leu Gly Leu Ala Tyr Phe Ala
                775                 780                 785

tca gtg ctc acc acc atc ggg aac tgg ctg cga gta gtg tcc cgc cgc      2457
Ser Val Leu Thr Thr Ile Gly Asn Trp Leu Arg Val Val Ser Arg Arg
            790                 795                 800

act cgg gca gag atg ggc ggc ctc acg gct cag gct gcc agc tgg act      2505
Thr Arg Ala Glu Met Gly Gly Leu Thr Ala Gln Ala Ala Ser Trp Thr
        805                 810                 815

ggc aca gtg aca gcg cgc gtg acc cag cga gcc ggg ccc gcc gcc ccg      2553
Gly Thr Val Thr Ala Arg Val Thr Gln Arg Ala Gly Pro Ala Ala Pro
    820                 825                 830

ccg ccg gag aag gag cag cca ctg ctg cct cca ccg ccc tgt cca gcg      2601
Pro Pro Glu Lys Glu Gln Pro Leu Leu Pro Pro Pro Pro Cys Pro Ala
835                 840                 845                 850

cag ccg ctg ggc agg ccc cga tcc cct tcg ccc ccc gag aag gct cag      2649
Gln Pro Leu Gly Arg Pro Arg Ser Pro Ser Pro Pro Glu Lys Ala Gln
                855                 860                 865
```

```
ccg cct tcc ccg ccc acg gcc tcg gcc ctg gat tat ccc agc gag aac      2697
Pro Pro Ser Pro Pro Thr Ala Ser Ala Leu Asp Tyr Pro Ser Glu Asn
            870                 875                 880

ctg gcc ttc atc gac gag tcc tcg gat acg cag agc gag cgc ggc tgc      2745
Leu Ala Phe Ile Asp Glu Ser Ser Asp Thr Gln Ser Glu Arg Gly Cys
        885                 890                 895

ccg ctg ccc cgc gcg ccg aga ggt cgc cgc cgc cca aat ccc ccc agg      2793
Pro Leu Pro Arg Ala Pro Arg Gly Arg Arg Arg Pro Asn Pro Pro Arg
    900                 905                 910

aag ccc gtg cgg ccc cgc ggc ccc ggg cgt ccc cga gac aaa ggc gtg      2841
Lys Pro Val Arg Pro Arg Gly Pro Gly Arg Pro Arg Asp Lys Gly Val
915                 920                 925                 930

ccg acc ccc cca agg ctt tct gtg tcg ctg ccc cgg gcg ggt gta tcc      2889
Pro Thr Pro Pro Arg Leu Ser Val Ser Leu Pro Arg Ala Gly Val Ser
                935                 940                 945

ctc aca gca cct cac gac tgt gcc tca aag cct gca tca ata aat gaa      2937
Leu Thr Ala Pro His Asp Cys Ala Ser Lys Pro Ala Ser Ile Asn Glu
            950                 955                 960

aac ggt ctg cac cgc tgc ggg cgt gac gct ccc gga cgc gag tgg gtg      2985
Asn Gly Leu His Arg Cys Gly Arg Asp Ala Pro Gly Arg Glu Trp Val
        965                 970                 975

tgg aat tgc ttt cct cgg gcc acc gtg ggg gca cct ctg gcc tcc cgt      3033
Trp Asn Cys Phe Pro Arg Ala Thr Val Gly Ala Pro Leu Ala Ser Arg
    980                 985                 990

gac ccc cag gcc gag ggt  ccc cgg gca ccc agc  ctt ggc tgc ccg        3078
Asp Pro Gln Ala Glu Gly  Pro Arg Ala Pro Ser  Leu Gly Cys Pro
995                 1000                 1005

cag  ccc cca ccc aac ccc  acg ttc tac ggg atc  ccc aac ccg gcc       3123
Gln  Pro Pro Pro Asn Pro  Thr Phe Tyr Gly Ile  Pro Asn Pro Ala
1010                 1015                 1020

cgg  ctc agt tcc cca gcc  cgc tct tcc ttc ccg  ctc cag cca tcc       3168
Arg  Leu Ser Ser Pro Ala  Arg Ser Ser Phe Pro  Leu Gln Pro Ser
1025                 1030                 1035

gcg  acc ctt ggc tcc ctc  ctt gta tgt ggc cca  cag gtg tcg ctc       3213
Ala  Thr Leu Gly Ser Leu  Leu Val Cys Gly Pro  Gln Val Ser Leu
1040                 1045                 1050

aag  tct tcc gac cgc caa  ggc tcg gac gag gag  agc gtg cat agc       3258
Lys  Ser Ser Asp Arg Gln  Gly Ser Asp Glu Glu  Ser Val His Ser
1055                 1060                 1065

gac  act cgg gac ctg tgg  acc acg acc acg ctg  tcc cag gca cag       3303
Asp  Thr Arg Asp Leu Trp  Thr Thr Thr Thr Leu  Ser Gln Ala Gln
1070                 1075                 1080

ctg  aac atg ccg ctg tcc  gag gtc tgc gag ggc  ttc gac gat gag       3348
Leu  Asn Met Pro Leu Ser  Glu Val Cys Glu Gly  Phe Asp Asp Glu
1085                 1090                 1095

ggc  cgc aac att agc aag  acc cgc ggg tgg cac  agc ccg ggg cgg       3393
Gly  Arg Asn Ile Ser Lys  Thr Arg Gly Trp His  Ser Pro Gly Arg
1100                 1105                 1110
```

ggc tcg ttg gac gag ggg tac aag gcc agc cac aag ccg gag gaa 3438
Gly Ser Leu Asp Glu Gly Tyr Lys Ala Ser His Lys Pro Glu Glu
1115 1120 1125

ctg gac gag cac gcg ctg gtg gag ctg gag ttg cac cgc ggc agc 3483
Leu Asp Glu His Ala Leu Val Glu Leu Glu Leu His Arg Gly Ser
1130 1135 1140

tcc atg gaa atc aat ctg ggg gag aag gac act gca tcc cag atc 3528
Ser Met Glu Ile Asn Leu Gly Glu Lys Asp Thr Ala Ser Gln Ile
1145 1150 1155

gag gcc gaa aag tct tcc tca atg tca tca ctc aat att gcg aag 3573
Glu Ala Glu Lys Ser Ser Ser Met Ser Ser Leu Asn Ile Ala Lys
1160 1165 1170

cac atg ccc cat cga gcc tac tgg gca gag cag cag agc agg gtt 3618
His Met Pro His Arg Ala Tyr Trp Ala Glu Gln Gln Ser Arg Val
1175 1180 1185

gga ggg gct ggg gag act ggg cgt ttc ggt ggg ctg cca ctg ccc 3663
Gly Gly Ala Gly Glu Thr Gly Arg Phe Gly Gly Leu Pro Leu Pro
1190 1195 1200

ctg atg gaa ctc atg gag aat gaa gct ctg gaa atc ctc acc aaa 3708
Leu Met Glu Leu Met Glu Asn Glu Ala Leu Glu Ile Leu Thr Lys
1205 1210 1215

gcc ctc cgg agt aag ctc ccc gcc aac ccc caa gaa ctc cca cga 3753
Ala Leu Arg Ser Lys Leu Pro Ala Asn Pro Gln Glu Leu Pro Arg
1220 1225 1230

cag att ctg gtg gat ttt gca ggg ctg ggg ccc agg ggg aga tgc 3798
Gln Ile Leu Val Asp Phe Ala Gly Leu Gly Pro Arg Gly Arg Cys
1235 1240 1245

aaa gtt ccc cag gct aac aca gac ctg agt gcc ctg cgc tac tgc 3843
Lys Val Pro Gln Ala Asn Thr Asp Leu Ser Ala Leu Arg Tyr Cys
1250 1255 1260

tac ctc gaa tca tct gcg gtt cct cga atc aca cat gcg gcg ccc 3888
Tyr Leu Glu Ser Ser Ala Val Pro Arg Ile Thr His Ala Ala Pro
1265 1270 1275

cct ggc tac cag tta ggg atc ggc agg gac cac ttc ctg act aag 3933
Pro Gly Tyr Gln Leu Gly Ile Gly Arg Asp His Phe Leu Thr Lys
1280 1285 1290

gag ctg cag cga tac atc gaa ggg ctc aag aag cgc cgg agc aag 3978
Glu Leu Gln Arg Tyr Ile Glu Gly Leu Lys Lys Arg Arg Ser Lys
1295 1300 1305

agg ctg tac gtg aat taa 3996
Arg Leu Tyr Val Asn
1310

<210> 29
<211> 1314
<212> PRT
<213> Homo sapiens

<400> 29

Met Glu Ser Asn Leu Ser Gly Leu Val Pro Ala Ala Gly Leu Val Pro
1   5   10   15

Ala Leu Pro Pro Ala Val Thr Leu Gly Leu Thr Ala Ala Tyr Thr Thr
            20   25   30

Leu Tyr Ala Leu Leu Phe Phe Ser Val Tyr Ala Gln Leu Trp Leu Val
        35   40   45

Leu Leu Tyr Gly His Lys Arg Leu Ser Tyr Gln Thr Val Phe Leu Ala
    50   55   60

Leu Cys Leu Leu Trp Ala Ala Leu Arg Thr Thr Leu Phe Ser Phe Tyr
65   70   75   80

Phe Arg Asp Thr Pro Arg Ala Asn Arg Leu Gly Pro Leu Pro Phe Trp
                85   90   95

Leu Leu Tyr Cys Cys Pro Val Cys Leu Gln Phe Phe Thr Leu Thr Leu
            100   105   110

Met Asn Leu Tyr Phe Ala Gln Val Val Phe Lys Ala Lys Val Lys Arg
        115   120   125

Arg Pro Glu Met Ser Arg Gly Leu Leu Ala Val Arg Gly Ala Phe Val
    130   135   140

Gly Ala Ser Leu Leu Phe Leu Leu Val Asn Val Leu Cys Ala Val Leu
145   150   155   160

Ser His Arg Arg Arg Ala Gln Pro Trp Ala Leu Leu Leu Val Arg Val
                165   170   175

Leu Val Ser Asp Ser Leu Phe Val Ile Cys Ala Leu Ser Leu Ala Ala
            180   185   190

Cys Leu Cys Leu Val Ala Arg Arg Ala Pro Ser Thr Ser Ile Tyr Leu
        195   200   205

Glu Ala Lys Gly Thr Ser Val Cys Gln Ala Ala Ala Met Gly Gly Ala
    210   215   220

Met Val Leu Leu Tyr Ala Ser Arg Ala Cys Tyr Asn Leu Thr Ala Leu
225   230   235   240

Ala Leu Ala Pro Gln Ser Arg Leu Asp Thr Phe Asp Tyr Asp Trp Tyr

245 250 255

Asn Val Ser Asp Gln Ala Asp Leu Val Asn Asp Leu Gly Asn Lys Gly
260 265 270

Tyr Leu Val Phe Gly Leu Ile Leu Phe Val Trp Glu Leu Leu Pro Thr
275 280 285

Thr Leu Leu Val Gly Phe Phe Arg Val His Arg Pro Pro Gln Asp Leu
290 295 300

Ser Thr Ser His Ile Leu Asn Gly Gln Val Phe Ala Ser Arg Ser Tyr
305 310 315 320

Phe Phe Asp Arg Ala Gly His Cys Glu Asp Glu Gly Cys Ser Trp Glu
325 330 335

His Ser Arg Gly Glu Ser Thr Ser Ser Cys Asp Cys Gly Pro Gly His
340 345 350

Cys Pro Glu Thr Asp Pro Val Ser Leu Leu Gln Tyr Val Gly Gln Ser
355 360 365

Arg Leu Trp Glu Leu Asn Thr Gln Ala Pro Val Pro Leu Thr Leu Gly
370 375 380

Pro Cys Ala Lys Phe Val Cys Arg Phe Leu Pro Arg Ile Leu Gly Val
385 390 395 400

Val Ala Thr Pro Ser Ser Gly Arg Leu Leu Ala Ala Pro Val Ile Asp
405 410 415

Ser Gly Ala Gly Thr Pro Gln Gly Arg Leu Ala Gly Arg Gly Ala His
420 425 430

Leu Ser Arg Val Gly Ala Ser Gly Ser Gly Val Ala Ala Gly Pro Ala
435 440 445

Ala Arg His Ala Pro Arg Arg Arg Cys Ala Asp Ala Gly Glu Ala Val
450 455 460

Gly Ala Ser Cys Gly Arg Cys Ala Val Ala Leu Leu Ser Gly Val Cys
465 470 475 480

Thr Leu Val Ser Thr His Val Cys Val Gly Ser Gly Cys Pro Gly Ala
485 490 495

Ala Gly Thr Pro Met Gly Ala Gly Asp Ala Gly Ala Ser Ala Glu Ser
500 505 510

Ala Val Thr Thr Ala Pro Gln Glu Pro Pro Ala Arg Pro Leu Gln Ala
515 520 525

Gly Ser Gly Ala Gly Pro Ala Pro Gly Arg Ala Met Arg Ser Thr Thr
530 535 540

Leu Leu Ala Leu Leu Ala Leu Val Leu Leu Tyr Leu Val Ser Gly Ala
545 550 555 560

Leu Val Phe Arg Ala Leu Glu Gln Pro His Glu Gln Gln Ala Gln Arg
565 570 575

Glu Leu Gly Glu Val Arg Glu Lys Phe Leu Arg Ala His Pro Cys Val
580 585 590

Ser Asp Gln Glu Leu Gly Leu Leu Ile Lys Glu Val Ala Asp Ala Leu
595 600 605

Gly Gly Gly Ala Asp Pro Glu Thr Asn Ser Thr Ser Asn Ser Ser His
610 615 620

Ser Ala Trp Asp Leu Gly Ser Ala Phe Phe Phe Ser Gly Thr Ile Ile
625 630 635 640

Thr Thr Ile Gly Tyr Gly Asn Val Ala Leu Arg Thr Asp Ala Gly Arg
645 650 655

Leu Phe Cys Ile Phe Tyr Ala Leu Val Gly Ile Pro Leu Phe Gly Ile
660 665 670

Leu Leu Ala Gly Val Gly Asp Arg Leu Gly Ser Ser Leu Arg His Gly
675 680 685

Ile Gly His Ile Glu Ala Ile Phe Leu Lys Trp His Val Pro Pro Glu
690 695 700

Leu Val Arg Val Leu Ser Ala Met Leu Phe Leu Leu Ile Gly Cys Leu
705 710 715 720

Leu Phe Val Leu Thr Pro Thr Phe Val Phe Cys Tyr Met Glu Asp Trp
725 730 735

Ser Lys Leu Glu Ala Ile Tyr Phe Val Ile Val Thr Leu Thr Thr Val
740 745 750

Gly Phe Gly Asp Tyr Val Ala Gly Ala Asp Pro Arg Gln Asp Ser Pro
755 760 765

Ala Tyr Gln Pro Leu Val Trp Phe Trp Ile Leu Leu Gly Leu Ala Tyr
770 775 780

Phe Ala Ser Val Leu Thr Thr Ile Gly Asn Trp Leu Arg Val Val Ser
785 790 795 800

Arg Arg Thr Arg Ala Glu Met Gly Gly Leu Thr Ala Gln Ala Ala Ser
805 810 815

Trp Thr Gly Thr Val Thr Ala Arg Val Thr Gln Arg Ala Gly Pro Ala
820 825 830

Ala Pro Pro Pro Glu Lys Glu Gln Pro Leu Leu Pro Pro Pro Pro Cys
835 840 845

Pro Ala Gln Pro Leu Gly Arg Pro Arg Ser Pro Ser Pro Pro Glu Lys
850 855 860

Ala Gln Pro Pro Ser Pro Pro Thr Ala Ser Ala Leu Asp Tyr Pro Ser
865 870 875 880

Glu Asn Leu Ala Phe Ile Asp Glu Ser Ser Asp Thr Gln Ser Glu Arg
885 890 895

Gly Cys Pro Leu Pro Arg Ala Pro Arg Gly Arg Arg Arg Pro Asn Pro
900 905 910

Pro Arg Lys Pro Val Arg Pro Arg Gly Pro Gly Arg Pro Arg Asp Lys
915 920 925

Gly Val Pro Thr Pro Pro Arg Leu Ser Val Ser Leu Pro Arg Ala Gly
930 935 940

Val Ser Leu Thr Ala Pro His Asp Cys Ala Ser Lys Pro Ala Ser Ile
945 950 955 960

Asn Glu Asn Gly Leu His Arg Cys Gly Arg Asp Ala Pro Gly Arg Glu
965 970 975

Trp Val Trp Asn Cys Phe Pro Arg Ala Thr Val Gly Ala Pro Leu Ala
980 985 990

Ser Arg Asp Pro Gln Ala Glu Gly Pro Arg Ala Pro Ser Leu Gly Cys
995 1000 1005

Pro Gln Pro Pro Pro Asn Pro Thr Phe Tyr Gly Ile Pro Asn Pro
1010 1015 1020

Ala Arg Leu Ser Ser Pro Ala Arg Ser Ser Phe Pro Leu Gln Pro
1025 1030 1035

Ser Ala Thr Leu Gly Ser Leu Leu Val Cys Gly Pro Gln Val Ser
1040 1045 1050

Leu Lys Ser Ser Asp Arg Gln Gly Ser Asp Glu Glu Ser Val His
1055 1060 1065

Ser Asp Thr Arg Asp Leu Trp Thr Thr Thr Thr Leu Ser Gln Ala
1070 1075 1080

Gln Leu Asn Met Pro Leu Ser Glu Val Cys Glu Gly Phe Asp Asp
1085 1090 1095

Glu Gly Arg Asn Ile Ser Lys Thr Arg Gly Trp His Ser Pro Gly
1100 1105 1110

Arg Gly Ser Leu Asp Glu Gly Tyr Lys Ala Ser His Lys Pro Glu
1115 1120 1125

Glu Leu Asp Glu His Ala Leu Val Glu Leu Glu Leu His Arg Gly
1130 1135 1140

Ser Ser Met Glu Ile Asn Leu Gly Glu Lys Asp Thr Ala Ser Gln
1145 1150 1155

Ile Glu Ala Glu Lys Ser Ser Ser Met Ser Ser Leu Asn Ile Ala
1160 1165 1170

Lys His Met Pro His Arg Ala Tyr Trp Ala Glu Gln Gln Ser Arg
1175 1180 1185

Val Gly Gly Ala Gly Glu Thr Gly Arg Phe Gly Gly Leu Pro Leu
1190 1195 1200

Pro Leu Met Glu Leu Met Glu Asn Glu Ala Leu Glu Ile Leu Thr
1205 1210 1215

Lys Ala Leu Arg Ser Lys Leu Pro Ala Asn Pro Gln Glu Leu Pro
1220 1225 1230

Arg Gln Ile Leu Val Asp Phe Ala Gly Leu Gly Pro Arg Gly Arg

1235 1240 1245

Cys Lys Val Pro Gln Ala Asn Thr Asp Leu Ser Ala Leu Arg Tyr
1250 1255 1260

Cys Tyr Leu Glu Ser Ser Ala Val Pro Arg Ile Thr His Ala Ala
1265 1270 1275

Pro Pro Gly Tyr Gln Leu Gly Ile Gly Arg Asp His Phe Leu Thr
1280 1285 1290

Lys Glu Leu Gln Arg Tyr Ile Glu Gly Leu Lys Lys Arg Arg Ser
1295 1300 1305

Lys Arg Leu Tyr Val Asn
1310

<210> 30
<211> 3945
<212> DNA
<213> Homo sapiens

<400> 30
atggagagta acctgtctgg cctggtgcct gctgccgggc tggtgcctgc gctgccacct 60

gctgtgaccc tggggctgac agctgcctac accaccctgt atgccctgct cttcttctcc 120

gtctatgccc agctctggct ggtgcttctg tatgggcaca agcgtctcag ctatcagacg 180

gtgttcctgg ccctctgtct gctctgggcc gccttgcgta ccaccctctt ctccttctac 240

ttccgagata ctccccgcgc caaccgcctg ggcccttgc ccttctggct tctctactgc 300

tgccccgtct gcctgcagtt cttcaccttg acgcttatga acctctactt tgcccaggtg 360

gtgttcaagg ccaaggtgaa gcgtcggccg gagatgagcc gaggcttgct cgctgtccga 420

ggggcctttg tgggggcctc gctgctcttt ctgctggtga acgtgctgtg tgctgtgctc 480

tcccatcggc gccgggcaca gccctgggcc ctgctgcttg tccgcgtcct ggtgagcgac 540

tccctgttcg tcatctgcgc gctgtctctt gctgcctgcc tctgcctcgt cgccaggcgg 600

gcgccctcca ctagcatcta cctggaggcc aaggggacca gtgtgtgcca ggcggccgcg 660

atgggtggcg ccatggtcct gctctatgcc agccgggcct gctacaacct gacagcactg 720

gccttggccc cccagagccg gctggacacc ttcgattacg actggtacaa tgtgtctgac 780

caggcggacc tggtgaatga cctggggaac aaaggctacc tggtatttgg cctcatcctc 840

ttcgtgtggg agctactgcc caccaccctg ctggtgggct tcttccgggt gcaccggccc 900

ccacaggacc tgagcaccag ccacatcctc aatgggcagg tctttgcctc tcggtcctac 960

ttctttgacc gggctgggca ctgtgaagat gagggctgct cctgggagca cagccggggt 1020

gagagcacca gctcttgtga ctgtggccct ggtcactgtc ctgagactga ccctgtttct 1080

ctgctgcagt atgtcgggca gtctaggctc tgggagctga atacccaggc cccagtcccc 1140

ctcaccctag gcccctgtgc caagtttgtc tgccgcttct tgcccaggat cctgggggtc 1200

gtggctaccc cctcctctgg ccggctcctt gctgctcctg tcatagactc aggggcaggg 1260

acgccccagg ggcgactggc cgggcggggt gcccacctct cgcgcgtggg cgcctccggg 1320

agtggtgtgg ccgccggtcc cgccgcccgc cacgctccga ggcgtcgctg tgcggacgcg 1380

ggggaggcgg tgggagcgag ctgcgggcgc tgcgcggtgg ccctgctgtc tggcgtgtgc 1440

acgctagtgt ccacacacgt gtgcgtgggc tctgggtgcc ctggcgcggc cggcacgccc 1500

atgggggccg gggatgccgg ggcgtctgcg gagagtgcag tgacgacagc tccccaggag 1560

ccccccgccc ggcccctcca ggcgggcagt ggagctggcc cggcgcctgg gcgcgccatg 1620

cgcagcacca cgctcctggc cctgctggcg ctggtcttgc tttacttggt gtctggtgcc 1680

ctggtgttcc gggccctgga gcagccccac gagcagcagg cccagaggga gctgggggag 1740

gtccgagaga agttcctgag ggcccatccg tgtgtgagcg accaggagct gggcctcctc 1800

atcaaggagg tggctgatgc cctgggaggg ggtgcggacc cagaaaccaa ctcgaccagc 1860

aacagcagcc actcagcctg ggacctgggc agcgccttct tttctcagg gaccatcatc 1920

accaccatcg gctatggcaa tgtggccctg cgcacagatg ccgggcgcct cttctgcatc 1980

ttttatgcgc tggtggggat tccgctgttt gggatcctac tggcaggggt cggggaccgg 2040

ctgggctcct ccctgcgcca tggcatcggt cacattgaag ccatcttctt gaagtggcac 2100

gtgccaccgg agctagtaag agtgctgtcg gcgatgcttt tcctgctgat cggctgcctg 2160

ctctttgtcc tcacgcccac gttcgtgttc tgctatatgg aggactggag caagctggag 2220

gccatctact ttgtcatagt gacgcttacc accgtgggct ttggcgacta tgtggccggc 2280

gcggacccca ggcaggactc ccggcctat cagccgctgg tgtggttctg gatcctgctc 2340

ggcctggctt acttcgcctc agtgctcacc accatcggga actggctgcg agtagtgtcc 2400

cgccgcactc gggcagagat gggcggcctc acggctcagg ctgccagctg gactggcaca 2460

gtgacagcgc gcgtgaccca gcgagccggg cccgccgccc cgccgccgga gaaggagcag 2520

ccactgctgc ctccaccgcc ctgtccagcg cagccgctgg gcaggccccg atccccttcg 2580

ccccccgaga aggctcagcc gccttccccg cccacggcct cggccctgga ttatcccagc 2640

gagaacctgg ccttcatcga cgagtcctcg gatacgcaga gcgagcgcgg ctgcccgctg 2700

ccccgcgcgc cgagaggtcg ccgccgccca aatcccccca ggaagcccgt gcggccccgc 2760

ggccccgggc gtccccgaga caaaggcgtg ccgacccccc caaggctttc tgtgtcgctg 2820

ccccgggcgg gtgtatccct cacagcacct cacgactgtg cctcaaagcc tgcatcaata 2880

aatgaaaacg gtctgcaccg ctgcgggcgt gacgctcccg gacgcgagtg ggtgtggaat 2940

tgctttcctc gggccaccgt gggggcacct ctggcctccc gtgaccccca ggccgagggt 3000

ccccgggcac ccagccttgg ctgcccgcag cccccaccca accccacgtt ctacgggatc 3060

cccaacccgg cccggctcag ttccccagcc cgctcttcct tcccgctcca gccatccgcg 3120

acccttggct ccctccttgt atgtggccca caggtgtcgc tcaagtcttc cgaccgccaa 3180

ggctcggacg aggagagcgt gcatagcgac actcgggacc tgtggaccac gaccacgctg 3240

tcccaggcac agctgaacat gccgctgtcc gaggtctgcg agggcttcga cgatgagggc 3300

cgcaacatta gcaagacccg cgggtggcac agcccggggc ggggctcgtt ggacgagggg 3360

tacaaggcca gccacaagcc ggaggaactg gacgagcacg cgctggtgga gctggagttg 3420

caccgcggca gctccatgga aatcaatctg gggagaagg acactgcatc ccagatcgag 3480

gccgaaaagt cttcctcaat gtcatcactc aatattgcga agcacatgcc ccatcgagcc 3540

tactgggcag agcagcagag cagggttgga ggggctgggg agactgggcg tttcggtggg 3600

ctgccactgc ccctgatgga actcatggag aatgaagctc tggaaatcct caccaaagcc 3660

ctccggagta agctccccgc caaccccaa gaactcccac gacagattct ggtggatttt 3720

gcagggctgg ggcccagggg gagatgcaaa gttccccagg ctaacacaga cctgagtgcc 3780

ctgcgctact gctacctcga atcatctgcg gttcctcgaa tcacacatgc ggcgccccct 3840

ggctaccagt tagggatcgg cagggaccac ttcctgacta aggagctgca gcgatacatc 3900

gaagggctca agaagcgccg gagcaagagg ctgtacgtga attaa 3945

<210> 31
<211> 178
<212> PRT
<213> Homo sapiens

<400> 31

Leu Ser Tyr Gln Thr Val Phe Leu Ala Leu Cys Leu Leu Trp Ala Ala
1 5 10 15

Leu Arg Thr Thr Leu Phe Ser Phe Tyr Phe Arg Asp Thr Pro Arg Ala
20 25 30

Asn Arg Leu Gly Pro Leu Pro Phe Trp Leu Leu Tyr Cys Cys Pro Val
35 40 45

Cys Leu Gln Phe Phe Thr Leu Thr Leu Met Asn Leu Tyr Phe Ala Gln
50 55 60

Val Val Phe Lys Ala Lys Val Lys Arg Arg Pro Glu Met Ser Arg Gly
65 70 75 80

Leu Leu Ala Val Arg Gly Ala Phe Val Gly Ala Ser Leu Leu Phe Leu
85 90 95

Leu Val Asn Val Leu Cys Ala Val Leu Ser His Arg Arg Arg Ala Gln
100 105 110

Pro Trp Ala Leu Leu Leu Val Arg Val Leu Val Ser Asp Ser Leu Phe
115 120 125

Val Ile Cys Ala Leu Ser Leu Ala Ala Cys Leu Cys Leu Val Ala Arg
130 135 140

Arg Ala Pro Ser Thr Ser Ile Tyr Leu Glu Ala Lys Gly Thr Ser Val
145 150 155 160

Cys Gln Ala Ala Ala Met Gly Gly Ala Met Val Leu Leu Tyr Ala Ser
165 170 175

Arg Ala

<210> 32
<211> 334
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(334)
<223> X = any amino acid or a stop codon

<400> 32

Val Arg Gly Leu Gly Pro Arg Leu Pro Val Phe Pro Lys Gly Lys Gly
1 5 10 15

Leu Ser Val Glu Glu Gly Gly Leu Ser Ala Thr Thr Ser Phe Leu Leu
20 25 30

Ser Ala Pro Ser Pro Ser Leu His Pro Ala Ile Pro Thr Pro Arg Ile
35 40 45

Tyr Phe Pro Gly Pro Ala Asp Ser Pro Ser Leu Ser Val Ser Arg Asp
50 55 60

Ser Gly Leu Pro Pro Leu Thr Trp Arg Val Thr Cys Leu Gly Leu Val
65 70 75 80

Ala Cys Leu Pro Gly Leu Val Pro Ala Leu Pro Pro Ala Val Thr Leu
85 90 95

Gly Leu Thr Ala Ala Tyr Thr Thr Leu Tyr Ala Leu Leu Phe Phe Ser
100 105 110

Val Tyr Ala Gln Leu Trp Leu Val Leu Arg Met Gly His Lys Arg Leu
115 120 125

Ser Tyr Gln Thr Val Phe Leu Ala Leu Cys Leu Phe Trp Ala Pro Leu
130 135 140

Arg Thr Thr Phe Phe Ser Phe Xaa Phe Pro Lys Ile Leu Pro Ala Pro
145 150 155 160

Asn Asn Ser Trp Gly Pro Leu Pro Phe Trp Leu Leu Tyr Cys Cys Pro
165 170 175

Val Cys Leu Gln Phe Phe Thr Leu Thr Leu Met Asn Leu Tyr Phe Ala

180 185 190

Gln Val Val Phe Lys Ala Lys Ser Glu Ala Ser Gly Pro Lys Met Ser
195 200 205

Arg Gly Leu Leu Ala Val Arg Gly Ala Phe Val Gly Ala Ser Leu Leu
210 215 220

Phe Leu Leu Val Asn Val Leu Cys Ala Val Leu Val Pro Cys Gly Ala
225 230 235 240

Ala Ala Gln Pro Trp Ala Leu Leu Leu Val Arg Val Leu Val Ser Asp
245 250 255

Ser Leu Phe Val Ile Cys Ala Leu Ser Leu Ala Ala Cys Leu Phe Leu
260 265 270

Cys Arg Gln Ala Gly Ala Leu His Xaa His Leu Pro Gly Gly Gln Gly
275 280 285

Arg Ala Ala Ala Leu Met Pro Arg Cys Leu Leu Gly Leu Ser Ala Ala
290 295 300

Val Leu Arg Val Xaa Arg Thr Ala Ala Glu Arg Pro Lys Arg His Leu
305 310 315 320

Gly Ile Ser Ala Ala Ala Leu Pro Trp Pro Pro Gly Arg Cys
325 330

<210> 33
<211> 443
<212> DNA
<213> Homo sapiens

<400> 33
cccacgcgtc cgatgaaatc agagcaagag gggctaatta gagattttac tagctagggc 60

tatttaccca aaccatctat ccagactgta gacatagaat caccagatgg ctaggatcct 120

ggtgcagctg gttccccttt ttccttaccc tgaatgtcat taaggatgca ttgccaaatg 180

ctgcccctct ggcctgatga ctacactcca tattggtcac ctgccttctt tttcctacag 240

tcttcctcca gacaggcacg ccatacaacc gactacactt gggctcactg aatgaatcac 300

attcttctgc tgtgcctccc agagatttca tcaaagcacc cgcagtggcc ttctggaggc 360

tcctcaaact ctccactcaa tgtttcctga aagtgtcggg tttctcacct ctagattgtt 420

attctcatca gttacatgtg ggg 443

<210> 34
<211> 2735
<212> DNA
<213> Homo sapiens

<400> 34
ttagattcag ctttgggaat ttgttccagg aactgcatca tccagcccaa aagcaaaaga 60

ggaaaatgca taagccagga tacctaaatg ttcctttatt taaacttgaa aatgtgactg 120

caaaaggagt caagaagttt agtactcaac acatttcttc tttgtagaaa taactgctca 180

```
agaagataaa ttgttaccta acagttgatt gatgaatatg aaatcagagc aagaggggct    240
aattagagaa attactagct agggctattt acccaaacca tctatccaga ctgtagacat    300
agaatcacca gatggctagg atcctggtgc agctggttcc ccttttttcct taccctgaat   360
gtcattaagg atgcattgcc aaatgctgcc cctctggcct gatgactaca ctccatattg    420
gtcacctgcc ttctttttcc tacagtcttc ctccagacag gcacgccata caaccgacta    480
cacttgggct cactgaatga atcacattct tctgctgtgc ctcccagaga tttcatcaaa    540
gcacccgcag tggccttctg gaggctcctc aaactctcca ctcaatgttt cctgaaagtg    600
tcgtgtttct cacctctaga ttgttattct catcagttac atgtgggttt cacaaattta    660
tttctcagaa tgcaagtctg tctcttatat cctcgggaaa cacaccttta tatcccagtt    720
agtactgaca aaaattaaac tagggactgg ccaaaaacag tgcctttcct cactttaatc    780
tcactaaagt agataagact caagttattt tgttcttgca atggcattga caaatgtttg    840
caccaaaaac catgttgaag ttcattaagg aaactgtgat ccaagatcca aggtcaaaaa    900
aacaaattca tcaattcagc acaccaccaa ctcacaggct aagcatctta ctgctaattc    960
attgatgctg ccatttgtca agtgccaaat tgaattattg atttgtcaat aatttccttc   1020
cgttggttac ttatatagta tattgcaatt cttgttgctg aagtcagcta cactttttct   1080
atttgaaaaa caatttcttg catttgggat ttcaggtata gtgattgtta caaatatgaa   1140
ggacttgaat taacagcaag ttttcaagta aaactttact tatgtataac tgaatgagtt   1200
cttaaagaca tttactaaca attttccaca aactaaaaat ttataaaaca ataaataaaa   1260
tagactttaa aaaaaagcgt gtcacacagc tgcttgtttt ttgtttgttt ctttgtttgt   1320
tttttagtag tgaaatggtg aaaaatcaga caatggtcac agagttcctc ctactgggat   1380
ttctcctggg cccaaggatt cagatgctcc tctttgggct cttctccctg ttctatgtct   1440
tcaccctgct ggggaatggg accatcctgg ggctcatctc actggactcc agactccaca   1500
cccccatgta cttcttcctc tcacacctgg ccgtcgtcaa catcgcctat gcctgcaaca   1560
cagtgcccca gatgctggtg aacctcctgc atccagccaa gcccatctcc tttgctggct   1620
gcatgacata gacctttctc tttttgagtt ttgcacatac tgaatgcctc ctgttggtgc   1680
tgatgtccta cgatcggtac gtggccatct gccaccctct ccgatatttc atcatcatga   1740
cctggaaagt ctgcatcact ctggccatca cttcctggac atgtggctcc ctcctggcta   1800
tggtccatgt gagcctcatc ctaagactgc ccttttgtgg gcctcgtgaa atcaaccact   1860
tcttctgtga aatcctgtct gtcctcaggc tggcctgtgc tgatacctgg ctcaaccagg   1920
tggtcatctt tgcagcctgc atgttcatcc tggtgggacc actctgcctg gtgctggtct   1980
cctactcaca catcctggcg gccatcctga ggatccagtc tggggagggc cgcagaaagg   2040
```

```
ccttctccac ctgctcctcc cacctctgcg tagtgggact cttctttggc agcgccatcg    2100

tcatgtacat ggcccctaag tcccgccatc ctgaggagca gcagaaggtc ctttttctat    2160

tttacagttc tttcaacccg atgctaaacc ccctgattta caacctgagg aatgtagagg    2220

tcaagggtgc cctgaggaga gcactgtgca aggaaagtca ttcctaagag gtgtgacatt    2280

tgaactgcca gcctcagttg tcacgtggac tcttgatgcc caattattgc ctcaatccag    2340

aaaagtttac ttctctttat ctgtgcttta ctgacagaag ggcaagtctt ctctcgtttt    2400

ttgcagataa aattttagat gtgttgcatt cattgggttt ctatgagatg tggttttatc    2460

agacaatttt ttcttttatt tcacaattac tttaatatct ttgaccagca aatatctcca    2520

ccctccaggg agaggtagta gcttctaggg aaaccatctt ggagagggtc ctgtcttccc    2580

ctgaggtggg ctttgaatcc agcactcttc cccttggagg gtcacttgga accagctaac    2640

ttttcagggt cctttcttcc cagttctgct catacatctg tcatgtaaca ctttagtgtt    2700

ctatatttgc atagctgtat ccttccatta gtttg                               2735
```

<210> 35
<211> 1788
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (833)..(1417)

<400> 35

```
atttggccct cgaggccaag aattcggcac gagggatcca aggtcaaaaa aacaaattca      60

tcaattcagc acaccaccaa ctcacaggct aagcatctta ctgctaattc attgatgctg     120

ccatttgtca agtgccaaat tgaattattg atttgtcaat aatttccttc cgttggttac     180

ttatatagta tattgcaatt cttgttgctg aagtcagcta cactttttct atttgaaaaa     240

caatttcttg catttgggat ttcaggtata gtgattgtta caaatatgaa ggacttgaat     300

taacagcaag ttttcaagta aaactttact tatgtataac tgaatgagtt cttaaagaca     360

tttactaaca attttccaca aactaaaaat ttataaaaca ataaataaaa tagactttaa     420

aaaaaagcgt gtcacacagc tgcttgtttt ttgtttgttt ctttgtttgt tttttagtag     480

tgaaatggtg aaaaatcaga caatggtcac agagttcctc ctactgggat ttctcctggg     540

cccaaggatt cagatgctcc tctttgggct cttctccctg ttctatgtct tcaccctgct     600

ggggaatggg accatcctgg ggctcatctc actggactcc agactccaca cccccatgta     660

cttcttcctc tcacacctgg ccgtcgtcaa catcgcctat gcctgcaaca cagtgcccca     720

gatgctggtg aacctcctgc atccagccaa gcccatctcc tttgctggct gcatgacata     780

gacctttctc tttttgagtt ttgcacatac tgaatgcctc ctgttggtgc tg atg tcc     838
                                                          Met Ser
```

```
                                                           1

tac gat cgg tac gtg gcc atc tgc cac cct ctc cga tat ttc atc atc      886
Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Phe Ile Ile
        5                  10                  15

atg acc tgg aaa gtc tgc atc act ctg gcc atc act tcc tgg aca tgt      934
Met Thr Trp Lys Val Cys Ile Thr Leu Ala Ile Thr Ser Trp Thr Cys
    20                  25                  30

ggc tcc ctc ctg gct atg gtc cat gtg agc ctc atc cta aga ctg ccc      982
Gly Ser Leu Leu Ala Met Val His Val Ser Leu Ile Leu Arg Leu Pro
35                  40                  45                  50

ttt tgt ggg cct cgt gaa atc aac cac ttc ttc tgt gaa atc ctg tct     1030
Phe Cys Gly Pro Arg Glu Ile Asn His Phe Phe Cys Glu Ile Leu Ser
                55                  60                  65

gtc ctc agg ctg gcc tgt gct gat acc tgg ctc aac cag gtg gtc atc     1078
Val Leu Arg Leu Ala Cys Ala Asp Thr Trp Leu Asn Gln Val Val Ile
            70                  75                  80

ttt gca gcc tgc atg ttc atc ctg gtg gga cca ctc tgc ctg gtg ctg     1126
Phe Ala Ala Cys Met Phe Ile Leu Val Gly Pro Leu Cys Leu Val Leu
        85                  90                  95

gtc tcc tac tca cac atc ctg gcg gcc atc ctg agg atc cag tct ggg     1174
Val Ser Tyr Ser His Ile Leu Ala Ala Ile Leu Arg Ile Gln Ser Gly
    100                 105                 110

gag ggc cgc aga aag gcc ttc tcc acc tgc tcc tcc cac ctc tgc gta     1222
Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser Ser His Leu Cys Val
115                 120                 125                 130

gtg gga ctc ttc ttt ggc agc gcc atc gtc atg tac atg gcc cct aag     1270
Val Gly Leu Phe Phe Gly Ser Ala Ile Val Met Tyr Met Ala Pro Lys
                135                 140                 145

tcc cgc cat cct gag gag cag cag aag gtc ctt ttt cta ttt tac agt     1318
Ser Arg His Pro Glu Glu Gln Gln Lys Val Leu Phe Leu Phe Tyr Ser
            150                 155                 160

tct ttc aac ccg atg cta aac ccc ctg att tac aac ctg agg aat gta     1366
Ser Phe Asn Pro Met Leu Asn Pro Leu Ile Tyr Asn Leu Arg Asn Val
        165                 170                 175

gag gtc aag ggt gcc ctg agg aga gca ctg tgc aag gaa agt cat tcc     1414
Glu Val Lys Gly Ala Leu Arg Arg Ala Leu Cys Lys Glu Ser His Ser
    180                 185                 190

taa gaggtgtgac atttgaactg ccagcctcag ttgtcacgtg gactcttgat          1467

gcccaattat tgcctcaatc cagaaaagtt tacttctctt tatctgtgct ttactgacag   1527

aagggcaagt cttctctcgt tttttgcaga taaaatttta gatgtgttgc attcattggg   1587

tttctatgag atgtggtttt atcagacaat tttttctttt atttcacaat tactttaata   1647

tctgtaaaat aaagaattat tttaattcat tttcccagtc ccaaaagtta aatacaggcc   1707

acttacttct ttaaccaaat gatatagttt ggctctgtgt ccccacccaa atctcatgtc   1767

aaattgtaat ccccgcatgt g                                             1788
```

<210> 36
<211> 194
<212> PRT
<213> Homo sapiens

<400> 36

Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Phe
1 5 10 15

Ile Ile Met Thr Trp Lys Val Cys Ile Thr Leu Ala Ile Thr Ser Trp
20 25 30

Thr Cys Gly Ser Leu Leu Ala Met Val His Val Ser Leu Ile Leu Arg
35 40 45

Leu Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe Phe Cys Glu Ile
50 55 60

Leu Ser Val Leu Arg Leu Ala Cys Ala Asp Thr Trp Leu Asn Gln Val
65 70 75 80

Val Ile Phe Ala Ala Cys Met Phe Ile Leu Val Gly Pro Leu Cys Leu
85 90 95

Val Leu Val Ser Tyr Ser His Ile Leu Ala Ala Ile Leu Arg Ile Gln
100 105 110

Ser Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser Ser His Leu
115 120 125

Cys Val Val Gly Leu Phe Phe Gly Ser Ala Ile Val Met Tyr Met Ala
130 135 140

Pro Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val Leu Phe Leu Phe
145 150 155 160

Tyr Ser Ser Phe Asn Pro Met Leu Asn Pro Leu Ile Tyr Asn Leu Arg
165 170 175

Asn Val Glu Val Lys Gly Ala Leu Arg Arg Ala Leu Cys Lys Glu Ser
180 185 190

His Ser

<210> 37
<211> 585

<212> DNA
<213> Homo sapiens

<400> 37
atgtcctacg atcggtacgt ggccatctgc caccctctcc gatatttcat catcatgacc 60

tggaaagtct gcatcactct ggccatcact tcctggacat gtggctccct cctggctatg 120

gtccatgtga gcctcatcct aagactgccc ttttgtgggc ctcgtgaaat caaccacttc 180

ttctgtgaaa tcctgtctgt cctcaggctg gcctgtgctg atacctggct caaccaggtg 240

gtcatctttg cagcctgcat gttcatcctg gtgggaccac tctgcctggt gctggtctcc 300

tactcacaca tcctggcggc catcctgagg atccagtctg gggagggccg cagaaaggcc 360

ttctccacct gctcctccca cctctgcgta gtgggactct tctttggcag cgccatcgtc 420

atgtacatgg cccctaagtc ccgccatcct gaggagcagc agaaggtcct tttctatttt 480

tacagttctt tcaacccgat gctaaacccc ctgatttaca acctgaggaa tgtagaggtc 540

aagggtgccc tgaggagagc actgtgcaag gaaagtcatt cctaa 585

<210> 38
<211> 173
<212> PRT
<213> Homo sapiens

<400> 38

Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Phe
1 5 10 15

Ile Ile Met Thr Trp Lys Val Cys Ile Thr Leu Ala Ile Thr Ser Trp
20 25 30

Thr Cys Gly Ser Leu Leu Ala Met Val His Val Ser Leu Ile Leu Arg
35 40 45

Leu Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe Phe Cys Glu Ile
50 55 60

Leu Ser Val Leu Arg Leu Ala Cys Ala Asp Thr Trp Leu Asn Gln Val
65 70 75 80

Val Ile Phe Ala Ala Cys Met Phe Ile Leu Val Gly Pro Leu Cys Leu
85 90 95

Val Leu Val Ser Tyr Ser His Ile Leu Ala Ala Ile Leu Arg Ile Gln
100 105 110

Ser Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser Ser His Leu
115 120 125

Cys Val Val Gly Leu Phe Phe Gly Ser Ala Ile Val Met Tyr Met Ala
130 135 140

Pro Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val Leu Phe Leu Phe
145 150 155 160

Tyr Ser Ser Phe Asn Pro Met Leu Asn Pro Leu Ile Tyr

165 170

<210> 39
<211> 35
<212> PRT
<213> Homo sapiens

<400> 39

Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Phe
1 5 10 15

Ile Ile Met Thr Trp Lys Val Cys Ile Thr Leu Ala Ile Thr Ser Trp
20 25 30

Thr Cys Gly
35

<210> 40
<211> 159
<212> PRT
<213> Homo sapiens

<400> 40

Ser Leu Leu Ala Met Val His Val Ser Leu Ile Leu Arg Leu Pro Phe
1 5 10 15

Cys Gly Pro Arg Glu Ile Asn His Phe Phe Cys Glu Ile Leu Ser Val
20 25 30

Leu Arg Leu Ala Cys Ala Asp Thr Trp Leu Asn Gln Val Val Ile Phe
35 40 45

Ala Ala Cys Met Phe Ile Leu Val Gly Pro Leu Cys Leu Val Leu Val
50 55 60

Ser Tyr Ser His Ile Leu Ala Ala Ile Leu Arg Ile Gln Ser Gly Glu
65 70 75 80

Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser Ser His Leu Cys Val Val
85 90 95

Gly Leu Phe Phe Gly Ser Ala Ile Val Met Tyr Met Ala Pro Lys Ser
100 105 110

Arg His Pro Glu Glu Gln Gln Lys Val Leu Phe Leu Phe Tyr Ser Ser
115 120 125

Phe Asn Pro Met Leu Asn Pro Leu Ile Tyr Asn Leu Arg Asn Val Glu
130 135 140

Val Lys Gly Ala Leu Arg Arg Ala Leu Cys Lys Glu Ser His Ser
145 150 155

<210> 41
<211> 1782
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (485)..(1411)

<400> 41

```
atttggccct cgaggccaag aattcggcac gagggatcca aggtcaaaaa aacaaattca      60

tcaattcagc acaccaccaa ctcacaggct aagcatctta ctgctaattc attgatgctg     120

ccatttgtca agtgccaaat tgaattattg atttgtcaat aatttccttc cgttggttac     180

ttatatagta tattgcaatt cttgttgctg aagtcagcta cactttttct atttgaaaaa     240

caatttcttg catttgggat ttcaggtata gtgattgtta caaatatgaa ggacttgaat     300

taacagcaag ttttcaagta aaactttact tatgtataac tgaatgagtt cttaaagaca     360

tttactaaca attttccaca aactaaaaat ttataaaaca ataaataaaa tagactttaa     420

aaaaaagcgt gtcacacagc tgcttgtttt ttgtttgttt ctttgtttgt tttttagtag     480

tgaa atg gtg aaa aat cag aca atg gtc aca gag ttc ctc cta ctg gga      529
     Met Val Lys Asn Gln Thr Met Val Thr Glu Phe Leu Leu Leu Gly
     1               5                   10                  15

ttt ctc ctg ggc cca agg att cag atg ctc ctc ttt ggg ctc ttc tcc       577
Phe Leu Leu Gly Pro Arg Ile Gln Met Leu Leu Phe Gly Leu Phe Ser
                20                  25                  30

ctg ttc tat atc ttc acc ctg ctg ggg aac ggg gcc atc ctg ggg ctc       625
Leu Phe Tyr Ile Phe Thr Leu Leu Gly Asn Gly Ala Ile Leu Gly Leu
            35                  40                  45

atc tca ctg gac tcc aga ctc cat acc ccc atg tac ttc ttc ctc tca       673
Ile Ser Leu Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ser
        50                  55                  60

cac ctg gct gtc gtc gac atc gcc tac acc cgc aac acg gtg ccc cag       721
His Leu Ala Val Val Asp Ile Ala Tyr Thr Arg Asn Thr Val Pro Gln
    65                  70                  75

atg ctg gcg aac ctc ctg cat cca gcc aag ccc atc tcc ttt gct ggt       769
Met Leu Ala Asn Leu Leu His Pro Ala Lys Pro Ile Ser Phe Ala Gly
80                  85                  90                  95

tgc atg acg cag acc ttt ctc tgt ttg agt ttt gga cac agc gaa tgt       817
Cys Met Thr Gln Thr Phe Leu Cys Leu Ser Phe Gly His Ser Glu Cys
                100                 105                 110

ctc ctg ctg gtg ctg atg tcc tac gat cgt tac gtg gcc atc tgc cac       865
Leu Leu Leu Val Leu Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His
            115                 120                 125

cct ctc cga tac tcc gtc atc atg acc tgc tgc atc act ctg gcc atc       913
Pro Leu Arg Tyr Ser Val Ile Met Thr Cys Cys Ile Thr Leu Ala Ile
        130                 135                 140

act tcc tgg aca tgt ggc tcc ctc ctg gct atg gtc cat gtg agc ctc       961
Thr Ser Trp Thr Cys Gly Ser Leu Leu Ala Met Val His Val Ser Leu
    145                 150                 155

atc cta aga ctg ccc ttt tgt ggg cct cgt gaa atc aac cac ttc ttc      1009
Ile Leu Arg Leu Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe Phe
160                 165                 170                 175

tgt gaa atc ctg tct gtc ctc agg ctg gcc tgt gct gat acc tgg ctc      1057
```

```
Cys Glu Ile Leu Ser Val Leu Arg Leu Ala Cys Ala Asp Thr Trp Leu
                180                 185                 190

aac cag gtg gtc atc ttt gca gcc tgc atg ttc atc ctg gtg gga cca      1105
Asn Gln Val Val Ile Phe Ala Ala Cys Met Phe Ile Leu Val Gly Pro
            195                 200                 205

ctc tgc ctg gtg ctg gtc tcc tac tca cac atc ctg gcg gcc atc ctg      1153
Leu Cys Leu Val Leu Val Ser Tyr Ser His Ile Leu Ala Ala Ile Leu
        210                 215                 220

agg atc cag tct ggg gag ggc cgc aga aag gcc ttc tcc acc tgc tcc      1201
Arg Ile Gln Ser Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser
    225                 230                 235

tcc cac ctc tgc gta gtg gga ctc ttc ttt ggc agc gcc atc gtc atg      1249
Ser His Leu Cys Val Val Gly Leu Phe Phe Gly Ser Ala Ile Val Met
240                 245                 250                 255

tac atg gcc cct aag tcc cgc cat cct gag gag cag cag aag gtc ctt      1297
Tyr Met Ala Pro Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val Leu
                260                 265                 270

ttt cta ttt tac agt tct ttc aac ccg atg cta aac ccc ctg att tac      1345
Phe Leu Phe Tyr Ser Ser Phe Asn Pro Met Leu Asn Pro Leu Ile Tyr
            275                 280                 285

aac ctg agg aat gta gag gtc aag ggt gcc ctg agg aga gca ctg tgc      1393
Asn Leu Arg Asn Val Glu Val Lys Gly Ala Leu Arg Arg Ala Leu Cys
        290                 295                 300

aag gaa agt cat tcc taa gaggtgtgac atttgaactg ccagcctcag             1441
Lys Glu Ser His Ser
    305

ttgtcacgtg gactcttgat gcccaattat tgcctcaatc cagaaaagtt tacttctctt    1501

tatctgtgct ttactgacag aagggcaagt cttctctcgt tttttgcaga taaaatttta    1561

gatgtgttgc attcattggg tttctatgag atgtggtttt atcagacaat tttttctttt    1621

atttcacaat tactttaata tctgtaaaat aaagaattat tttaattcat tttcccagtc    1681

ccaaaagtta aatacaggcc acttacttct ttaaccaaat gatatagttt ggctctgtgt    1741

ccccacccaa atctcatgtc aaattgtaat ccccgcatgt g                        1782
```

<210> 42
<211> 308
<212> PRT
<213> Homo sapiens

<400> 42

```
Met Val Lys Asn Gln Thr Met Val Thr Glu Phe Leu Leu Leu Gly Phe
1               5                   10                  15

Leu Leu Gly Pro Arg Ile Gln Met Leu Leu Phe Gly Leu Phe Ser Leu
            20                  25                  30
```

Phe Tyr Ile Phe Thr Leu Leu Gly Asn Gly Ala Ile Leu Gly Leu Ile
35 40 45

Ser Leu Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ser His
50 55 60

Leu Ala Val Val Asp Ile Ala Tyr Thr Arg Asn Thr Val Pro Gln Met
65 70 75 80

Leu Ala Asn Leu Leu His Pro Ala Lys Pro Ile Ser Phe Ala Gly Cys
85 90 95

Met Thr Gln Thr Phe Leu Cys Leu Ser Phe Gly His Ser Glu Cys Leu
100 105 110

Leu Leu Val Leu Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro
115 120 125

Leu Arg Tyr Ser Val Ile Met Thr Cys Cys Ile Thr Leu Ala Ile Thr
130 135 140

Ser Trp Thr Cys Gly Ser Leu Leu Ala Met Val His Val Ser Leu Ile
145 150 155 160

Leu Arg Leu Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe Phe Cys
165 170 175

Glu Ile Leu Ser Val Leu Arg Leu Ala Cys Ala Asp Thr Trp Leu Asn
180 185 190

Gln Val Val Ile Phe Ala Ala Cys Met Phe Ile Leu Val Gly Pro Leu
195 200 205

Cys Leu Val Leu Val Ser Tyr Ser His Ile Leu Ala Ala Ile Leu Arg
210 215 220

Ile Gln Ser Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser Ser
225 230 235 240

His Leu Cys Val Val Gly Leu Phe Phe Gly Ser Ala Ile Val Met Tyr
245 250 255

Met Ala Pro Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val Leu Phe
260 265 270

Leu Phe Tyr Ser Ser Phe Asn Pro Met Leu Asn Pro Leu Ile Tyr Asn
275 280 285

Leu Arg Asn Val Glu Val Lys Gly Ala Leu Arg Arg Ala Leu Cys Lys
290 295 300

Glu Ser His Ser
305

<210> 43
<211> 927
<212> DNA
<213> Homo sapiens

<400> 43
atggtgaaaa atcagacaat ggtcacagag ttcctcctac tgggatttct cctgggccca 60

aggattcaga tgctcctctt tgggctcttc tccctgttct atatcttcac cctgctgggg 120

aacggggcca tcctggggct catctcactg gactccagac tccatacccc catgtacttc 180

ttcctctcac acctggctgt cgtcgacatc gcctacaccc gcaacacggt gccccagatg 240

ctggcgaacc tcctgcatcc agccaagccc atctcctttg ctggttgcat gacgcagacc 300

tttctctgtt tgagttttgg acacagcgaa tgtctcctgc tggtgctgat gtcctacgat 360

cgttacgtgg ccatctgcca ccctctccga tactccgtca tcatgacctg ctgcatcact 420

ctggccatca cttcctggac atgtggctcc ctcctggcta tggtccatgt gagcctcatc 480

ctaagactgc ccttttgtgg gcctcgtgaa atcaaccact tcttctgtga aatcctgtct 540

gtcctcaggc tggcctgtgc tgatacctgg ctcaaccagg tggtcatctt tgcagcctgc 600

atgttcatcc tggtgggacc actctgcctg gtgctggtct cctactcaca catcctggcg 660

gccatcctga ggatccagtc tggggagggc cgcagaaagg ccttctccac ctgctcctcc 720

cacctctgcg tagtgggact cttctttggc agcgccatcg tcatgtacat ggcccctaag 780

tcccgccatc ctgaggagca gcagaaggtc ctttttctat tttacagttc tttcaacccg 840

atgctaaacc ccctgattta caacctgagg aatgtagagg tcaagggtgc cctgaggaga 900

gcactgtgca aggaaagtca ttcctaa 927

<210> 44
<211> 248
<212> PRT
<213> Homo sapiens

<400> 44

Gly Asn Gly Ala Ile Leu Gly Leu Ile Ser Leu Asp Ser Arg Leu His
1 5 10 15

Thr Pro Met Tyr Phe Phe Leu Ser His Leu Ala Val Val Asp Ile Ala
20 25 30

Tyr Thr Arg Asn Thr Val Pro Gln Met Leu Ala Asn Leu Leu His Pro
35 40 45

Ala Lys Pro Ile Ser Phe Ala Gly Cys Met Thr Gln Thr Phe Leu Cys
50 55 60

Leu Ser Phe Gly His Ser Glu Cys Leu Leu Leu Val Leu Met Ser Tyr
65 70 75 80

Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Ser Val Ile Met
85 90 95

Thr Cys Cys Ile Thr Leu Ala Ile Thr Ser Trp Thr Cys Gly Ser Leu
100 105 110

Leu Ala Met Val His Val Ser Leu Ile Leu Arg Leu Pro Phe Cys Gly
115 120 125

Pro Arg Glu Ile Asn His Phe Phe Cys Glu Ile Leu Ser Val Leu Arg
130 135 140

Leu Ala Cys Ala Asp Thr Trp Leu Asn Gln Val Val Ile Phe Ala Ala
145 150 155 160

Cys Met Phe Ile Leu Val Gly Pro Leu Cys Leu Val Leu Val Ser Tyr
165 170 175

Ser His Ile Leu Ala Ala Ile Leu Arg Ile Gln Ser Gly Glu Gly Arg
180 185 190

Arg Lys Ala Phe Ser Thr Cys Ser Ser His Leu Cys Val Val Gly Leu
195 200 205

Phe Phe Gly Ser Ala Ile Val Met Tyr Met Ala Pro Lys Ser Arg His
210 215 220

Pro Glu Glu Gln Gln Lys Val Leu Phe Leu Phe Tyr Ser Ser Phe Asn
225 230 235 240

Pro Met Leu Asn Pro Leu Ile Tyr
245

<210> 45
<211> 42
<212> PRT
<213> Homo sapiens

<400> 45

Met Val Lys Asn Gln Thr Met Val Thr Glu Phe Leu Leu Leu Gly Phe
1 5 10 15

Leu Leu Gly Pro Arg Ile Gln Met Leu Leu Phe Gly Leu Phe Ser Leu
20 25 30

Phe Tyr Ile Phe Thr Leu Leu Gly Asn Gly
35 40

<210> 46
<211> 266
<212> PRT
<213> Homo sapiens

<400> 46

Ala Ile Leu Gly Leu Ile Ser Leu Asp Ser Arg Leu His Thr Pro Met
1 5 10 15

Tyr Phe Phe Leu Ser His Leu Ala Val Val Asp Ile Ala Tyr Thr Arg
20 25 30

Asn Thr Val Pro Gln Met Leu Ala Asn Leu Leu His Pro Ala Lys Pro
35 40 45

Ile Ser Phe Ala Gly Cys Met Thr Gln Thr Phe Leu Cys Leu Ser Phe
50 55 60

Gly His Ser Glu Cys Leu Leu Leu Val Leu Met Ser Tyr Asp Arg Tyr
65 70 75 80

Val Ala Ile Cys His Pro Leu Arg Tyr Ser Val Ile Met Thr Cys Cys
85 90 95

Ile Thr Leu Ala Ile Thr Ser Trp Thr Cys Gly Ser Leu Leu Ala Met
100 105 110

Val His Val Ser Leu Ile Leu Arg Leu Pro Phe Cys Gly Pro Arg Glu
115 120 125

Ile Asn His Phe Phe Cys Glu Ile Leu Ser Val Leu Arg Leu Ala Cys
130 135 140

Ala Asp Thr Trp Leu Asn Gln Val Val Ile Phe Ala Ala Cys Met Phe
145 150 155 160

Ile Leu Val Gly Pro Leu Cys Leu Val Leu Val Ser Tyr Ser His Ile
165 170 175

Leu Ala Ala Ile Leu Arg Ile Gln Ser Gly Glu Gly Arg Arg Lys Ala
180 185 190

Phe Ser Thr Cys Ser Ser His Leu Cys Val Val Gly Leu Phe Phe Gly
195 200 205

Ser Ala Ile Val Met Tyr Met Ala Pro Lys Ser Arg His Pro Glu Glu
210 215 220

Gln Gln Lys Val Leu Phe Leu Phe Tyr Ser Ser Phe Asn Pro Met Leu
225 230 235 240

Asn Pro Leu Ile Tyr Asn Leu Arg Asn Val Glu Val Lys Gly Ala Leu
245 250 255

Arg Arg Ala Leu Cys Lys Glu Ser His Ser
260 265

<210> 47
<211> 353
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (1)..(353)
<223> X = any amino acid or a stop codon

<400> 47

Arg His Leu Leu Thr Ile Phe His Lys Leu Lys Ile Tyr Lys Thr Ile
1 5 10 15

Asn Lys Ile Asp Phe Lys Lys Lys Arg Val Thr Gln Leu Leu Val Phe
20 25 30

Cys Leu Phe Leu Cys Leu Phe Phe Ser Ser Glu Met Val Lys Asn Gln
35 40 45

Thr Met Val Thr Glu Phe Leu Leu Leu Gly Phe Leu Leu Gly Pro Arg
50 55 60

Ile Gln Met Leu Leu Phe Gly Leu Phe Ser Leu Phe Tyr Val Phe Thr
65 70 75 80

Leu Leu Gly Asn Gly Thr Ile Leu Gly Leu Ile Ser Leu Asp Ser Arg
85 90 95

Leu His Thr Pro Met Tyr Phe Phe Leu Ser His Leu Ala Val Val Asn
100 105 110

Ile Ala Tyr Ala Cys Asn Thr Val Pro Gln Met Leu Val Asn Leu Leu
115 120 125

His Pro Ala Lys Pro Ile Ser Phe Ala Gly Cys Met Thr Xaa Thr Phe
130 135 140

Leu Phe Leu Ser Phe Ala His Thr Glu Cys Leu Leu Leu Val Leu Met
145 150 155 160

Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Phe Ile
165 170 175

Ile Met Thr Trp Lys Val Cys Ile Thr Leu Ala Ile Thr Ser Trp Thr
180 185 190

Cys Gly Ser Leu Leu Ala Met Val His Val Ser Leu Ile Leu Arg Leu
195 200 205

Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe Phe Cys Glu Ile Leu
210 215 220

Ser Val Leu Arg Leu Ala Cys Ala Asp Thr Trp Leu Asn Gln Val Val
225 230 235 240

Ile Phe Ala Ala Cys Met Phe Ile Leu Val Gly Pro Leu Cys Leu Val
245 250 255

Leu Val Ser Tyr Ser His Ile Leu Ala Ala Ile Leu Arg Ile Gln Ser
260 265 270

Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser Ser His Leu Cys
275 280 285

Val Val Gly Leu Phe Phe Gly Ser Ala Ile Val Met Tyr Met Ala Pro
290 295 300

Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val Leu Phe Leu Phe Tyr
305 310 315 320

Ser Ser Phe Asn Pro Met Leu Asn Pro Leu Ile Tyr Asn Leu Arg Asn
325 330 335

Val Glu Val Lys Gly Ala Leu Arg Arg Ala Leu Cys Lys Glu Ser His
340 345 350

Ser

<210> 48
<211> 517
<212> PRT
<213> Homo sapiens

<400> 48

Glu Asn Trp Arg Gln Lys Lys Lys Thr Leu Leu Val Ala Ile Asp Arg
1 5 10 15

Ala Cys Pro Glu Ser Gly His Pro Arg Val Leu Ala Asp Ser Phe Pro
20 25 30

Gly Ser Ser Pro Tyr Glu Gly Tyr Asn Tyr Gly Ser Phe Glu Asn Val
35 40 45

Ser Gly Ser Thr Asp Gly Leu Val Asp Ser Ala Gly Thr Gly Asp Leu
50 55 60

Ser Tyr Gly Tyr Gln Gly His Asp Gln Phe Lys Arg Arg Leu Pro Ser
65 70 75 80

Gly Gln Met Arg Gln Leu Cys Ile Ala Met Gly Arg Ser Phe Glu Pro
85 90 95

Val Gly Thr Arg Pro Arg Val Asp Ser Met Ser Ser Val Glu Glu Asp
100 105 110

Asp Tyr Asp Thr Leu Thr Asp Ile Asp Ser Asp Lys Asn Val Ile Arg
115 120 125

Thr Lys Gln Tyr Leu Tyr Val Ala Asp Leu Ala Arg Lys Asp Lys Arg
130 135 140

Val Leu Arg Lys Lys Tyr Gln Ile Tyr Phe Trp Asn Ile Ala Thr Ile
145 150 155 160

Ala Val Phe Tyr Ala Leu Pro Val Val Gln Leu Val Ile Thr Tyr Gln
165 170 175

Thr Val Val Asn Val Thr Gly Asn Gln Asp Ile Cys Tyr Tyr Asn Phe
180 185 190

Leu Cys Ala His Pro Leu Gly Asn Leu Ser Ala Phe Asn Asn Ile Leu
195 200 205

Ser Asn Leu Gly Tyr Ile Leu Leu Gly Leu Leu Phe Leu Leu Ile Ile
210 215 220

Leu Gln Arg Glu Ile Asn His Asn Arg Ala Leu Leu Arg Asn Asp Leu
225 230 235 240

Cys Ala Leu Glu Cys Gly Ile Pro Lys His Phe Gly Leu Phe Tyr Ala
245 250 255

Met Gly Thr Ala Leu Met Met Glu Gly Leu Leu Ser Ala Cys Tyr His

260 265 270

Val Cys Pro Asn Tyr Thr Asn Phe Gln Phe Asp Thr Ser Phe Met Tyr
275 280 285

Met Ile Ala Gly Leu Cys Met Leu Lys Leu Tyr Gln Lys Arg His Pro
290 295 300

Asp Ile Asn Ala Ser Ala Tyr Ser Ala Tyr Ala Cys Leu Ala Ile Val
305 310 315 320

Ile Phe Phe Ser Val Leu Gly Val Val Phe Gly Lys Gly Asn Thr Ala
325 330 335

Phe Trp Ile Val Phe Ser Ile Ile His Ile Ile Ala Thr Leu Leu Leu
340 345 350

Ser Thr Gln Leu Tyr Tyr Met Gly Arg Trp Lys Leu Asp Ser Gly Ile
355 360 365

Phe Arg Arg Ile Leu His Val Leu Tyr Thr Asp Cys Ile Arg Gln Cys
370 375 380

Ser Gly Pro Leu Tyr Val Asp Arg Met Val Leu Leu Val Met Gly Asn
385 390 395 400

Val Ile Asn Trp Ser Leu Ala Ala Tyr Gly Leu Ile Met Arg Pro Asn
405 410 415

Asp Phe Ala Ser Tyr Leu Leu Ala Ile Gly Ile Cys Asn Leu Leu Leu
420 425 430

Tyr Phe Ala Phe Tyr Ile Ile Met Lys Leu Arg Ser Gly Glu Arg Ile
435 440 445

Lys Leu Ile Pro Leu Leu Cys Ile Val Cys Thr Ser Val Val Trp Gly
450 455 460

Phe Ala Leu Phe Phe Phe Phe Gln Gly Leu Ser Thr Trp Gln Lys Thr
465 470 475 480

Pro Ala Glu Ser Arg Glu His Asn Arg Asp Cys Ile Leu Leu Asp Phe
485 490 495

Phe Asp Asp His Asp Ile Trp His Phe Leu Ser Ser Ile Ala Met Phe
500 505 510

Gly Ser Phe Leu Val
515

<210> 49
<211> 444
<212> PRT
<213> Homo sapiens

<400> 49

Gly His Arg Ala Ser Gln Thr Gln Thr Ala Pro Val Glu Glu Ser Asp
1 5 10 15

Phe Asp Thr Met Pro Asp Ile Glu Ser Asp Lys Asn Ile Ile Arg Thr
20 25 30

Lys Met Phe Leu Tyr Leu Ser Asp Leu Ser Arg Lys Asp Arg Arg Ile
35 40 45

Val Ser Lys Lys Tyr Lys Ile Tyr Phe Trp Asn Ile Ile Thr Ile Ala
50 55 60

Val Phe Tyr Ala Leu Pro Val Ile Gln Leu Val Ile Thr Tyr Gln Thr
65 70 75 80

Val Val Asn Val Thr Gly Asn Gln Asp Ile Cys Tyr Tyr Asn Phe Leu
85 90 95

Cys Ala His Pro Leu Gly Val Leu Ser Ala Phe Asn Asn Ile Leu Ser
100 105 110

Asn Leu Gly His Val Leu Leu Gly Phe Leu Phe Leu Leu Ile Val Leu
115 120 125

Arg Arg Asp Ile Leu His Arg Arg Ala Leu Glu Ala Lys Asp Ile Phe
130 135 140

Ala Val Glu Tyr Gly Ile Pro Lys His Phe Gly Leu Phe Tyr Ala Met
145 150 155 160

Gly Ile Ala Leu Met Met Glu Gly Val Leu Ser Ala Cys Tyr His Val
165 170 175

Cys Pro Asn Tyr Ser Asn Phe Gln Phe Asp Thr Ser Phe Met Tyr Met
180 185 190

Ile Ala Gly Leu Cys Met Leu Lys Leu Tyr Gln Thr Arg His Pro Asp
195 200 205

Ile Asn Ala Ser Ala Tyr Ser Ala Tyr Ala Ser Phe Ala Val Val Ile
210 215 220

Met Val Thr Val Leu Gly Val Val Phe Gly Lys Asn Asp Val Trp Phe
225 230 235 240

Trp Val Ile Phe Ser Ala Ile His Val Leu Ala Ser Leu Ala Leu Ser
245 250 255

Thr Gln Ile Tyr Tyr Met Gly Arg Phe Lys Ile Asp Val Ser Asp Thr
260 265 270

Asp Leu Gly Ile Phe Arg Arg Ala Ala Met Val Phe Tyr Thr Asp Cys
275 280 285

Ile Gln Gln Cys Ser Arg Pro Leu Tyr Met Asp Arg Met Val Leu Leu
290 295 300

Val Val Gly Asn Leu Val Asn Trp Ser Phe Ala Leu Phe Gly Leu Ile
305 310 315 320

Tyr Arg Pro Arg Asp Phe Ala Ser Tyr Met Leu Gly Ile Phe Ile Cys
325 330 335

Asn Leu Leu Leu Tyr Leu Ala Phe Tyr Ile Ile Met Lys Leu Arg Ser
340 345 350

Ser Glu Lys Val Leu Pro Val Pro Leu Phe Cys Ile Val Ala Thr Ala
355 360 365

```
Val Met Trp Ala Ala Ala Leu Tyr Phe Phe Phe Gln Asn Leu Ser Ser
    370                 375                 380

Trp Glu Gly Thr Pro Ala Glu Ser Arg Glu Lys Asn Arg Glu Cys Ile
385                 390                 395                 400

Leu Leu Asp Phe Phe Asp Asp His Asp Ile Trp His Phe Leu Ser Ala
                405                 410                 415

Thr Ala Leu Phe Phe Ser Phe Leu Val Leu Leu Thr Leu Asp Asp Asp
            420                 425                 430

Leu Asp Val Val Arg Arg Asp Gln Ile Pro Val Phe
        435                 440
```

<210> 50
<211> 267
<212> PRT
<213> Homo sapiens

<400> 50

```
Leu Phe Pro Gln Trp His Leu Pro Ile Lys Ile Ala Ala Ile Ile Ala
1               5                   10                  15

Ser Leu Thr Phe Leu Tyr Thr Leu Leu Arg Glu Val Ile His Pro Leu
            20                  25                  30

Ala Thr Ser His Gln Gln Tyr Phe Tyr Lys Ile Pro Ile Leu Val Ile
        35                  40                  45

Asn Lys Val Leu Pro Met Val Ser Ile Thr Leu Leu Ala Leu Val Tyr
    50                  55                  60

Leu Pro Gly Val Ile Ala Ala Ile Val Gln Leu His Asn Gly Thr Lys
65                  70                  75                  80

Tyr Lys Lys Phe Pro His Trp Leu Asp Lys Trp Met Leu Thr Arg Lys
                85                  90                  95

Gln Phe Gly Leu Leu Ser Phe Phe Phe Ala Val Leu His Ala Ile Tyr
            100                 105                 110

Ser Leu Ser Tyr Pro Met Arg Arg Ser Tyr Arg Tyr Lys Leu Leu Asn
        115                 120                 125

Trp Ala Tyr Gln Gln Val Gln Gln Asn Lys Glu Asp Ala Trp Ile Glu
    130                 135                 140

His Asp Val Trp Arg Met Glu Ile Tyr Val Ser Leu Gly Ile Val Gly
145                 150                 155                 160

Leu Ala Ile Leu Ala Leu Leu Ala Val Thr Ser Ile Pro Ser Val Ser
                165                 170                 175

Asp Ser Leu Thr Trp Arg Glu Phe His Tyr Ile Gln Ser Lys Leu Gly
            180                 185                 190

Ile Val Ser Leu Leu Leu Gly Thr Ile His Ala Leu Ile Phe Ala Trp
        195                 200                 205

Asn Lys Trp Ile Asp Ile Lys Gln Phe Val Trp Tyr Thr Pro Pro Thr
    210                 215                 220
```

Phe Met Ile Ala Val Phe Leu Pro Ile Val Val Leu Ile Phe Lys Ser
225 230 235 240

Ile Leu Phe Leu Pro Cys Leu Arg Lys Lys Ile Leu Lys Ile Arg His
245 250 255

Gly Trp Glu Asp Val Thr Lys Ile Asn Lys Thr
260 265

<210> 51
<211> 267
<212> PRT
<213> Homo sapiens

<400> 51

Leu Phe Pro Gln Trp His Leu Pro Ile Lys Ile Ala Ala Ile Ile Ala
1 5 10 15

Ser Leu Thr Phe Leu Tyr Thr Leu Leu Arg Glu Val Ile His Pro Leu
20 25 30

Ala Thr Ser His Gln Gln Tyr Phe Tyr Lys Ile Pro Ile Leu Val Ile
35 40 45

Asn Lys Val Leu Pro Met Val Ser Ile Thr Leu Leu Ala Leu Val Tyr
50 55 60

Leu Pro Gly Val Ile Ala Ala Ile Val Gln Leu His Asn Gly Thr Lys
65 70 75 80

Tyr Lys Lys Phe Pro His Trp Leu Asp Lys Trp Met Leu Thr Arg Lys
85 90 95

Gln Phe Gly Leu Leu Ser Phe Phe Phe Ala Val Leu His Ala Ile Tyr
100 105 110

Ser Leu Ser Tyr Pro Met Arg Arg Ser Tyr Arg Tyr Lys Leu Leu Asn
115 120 125

Trp Ala Tyr Gln Gln Val Gln Gln Asn Lys Glu Asp Ala Trp Ile Glu
130 135 140

His Asp Val Trp Arg Met Glu Ile Tyr Val Ser Leu Gly Ile Val Gly
145 150 155 160

Leu Ala Ile Leu Ala Leu Leu Ala Val Thr Ser Ile Pro Ser Val Ser
165 170 175

Asp Ser Leu Thr Trp Arg Glu Phe His Tyr Ile Gln Ser Lys Leu Gly
180 185 190

Ile Val Ser Leu Leu Leu Gly Thr Ile His Ala Leu Ile Phe Ala Trp
195 200 205

Asn Lys Trp Ile Asp Ile Lys Gln Phe Val Trp Tyr Thr Pro Pro Thr
210 215 220

Phe Met Ile Ala Val Phe Leu Pro Ile Val Val Leu Ile Phe Lys Ser
225 230 235 240

Ile Leu Phe Leu Pro Cys Leu Arg Lys Lys Ile Leu Lys Ile Arg His

245 250 255

Gly Trp Glu Asp Val Thr Lys Ile Asn Lys Thr
260 265

<210> 52
<211> 1349
<212> PRT
<213> Rattus norvegicus

<400> 52

Met Lys Ser Ser Arg Thr Val Thr Leu Tyr Phe Val Leu Ile Val Ile
1 5 10 15

Cys Ser Ser Glu Ala Thr Trp Ser Arg Pro Ala Glu Pro Ile Val His
20 25 30

Pro Leu Ile Leu Gln Glu His Glu Leu Ala Gly Glu Glu Leu Leu Arg
35 40 45

Pro Lys Arg Ala Val Ala Val Gly Gly Pro Val Ala Glu Glu Tyr Thr
50 55 60

Val Asp Val Glu Ile Ser Phe Glu Asn Val Ser Phe Leu Glu Ser Ile
65 70 75 80

Arg Ala His Leu Asn Ser Leu Arg Phe Pro Val Gln Gly Asn Gly Thr
85 90 95

Asp Ile Leu Ser Met Ala Met Thr Thr Val Cys Thr Pro Thr Gly Asn
100 105 110

Asp Leu Leu Cys Phe Cys Glu Lys Gly Tyr Gln Trp Pro Glu Glu Arg
115 120 125

Cys Leu Ser Ser Leu Thr Cys Gln Glu His Asp Ser Ala Leu Pro Gly
130 135 140

Arg Tyr Cys Asn Cys Leu Lys Gly Leu Pro Pro Gln Gly Pro Phe Cys
145 150 155 160

Gln Leu Pro Glu Thr Tyr Ile Thr Leu Lys Ile Lys Val Arg Leu Asn
165 170 175

Ile Gly Phe Gln Glu Asp Leu Glu Asn Thr Ser Ser Ala Leu Tyr Arg
180 185 190

Ser Tyr Lys Thr Asp Leu Glu Arg Ala Phe Arg Ala Gly Tyr Arg Thr
195 200 205

Leu Pro Gly Phe Arg Ser Val Thr Val Thr Gln Phe Thr Lys Gly Ser
210 215 220

Val Val Val Asp Tyr Ile Val Glu Val Ala Ser Ala Pro Leu Pro Gly
225 230 235 240

Ser Ile His Lys Ala Asn Glu Gln Val Ile Gln Asn Leu Asn Gln Thr
245 250 255

Tyr Lys Met Asp Tyr Asn Ser Phe Gln Gly Thr Pro Ser Asn Glu Thr
260 265 270

Lys Phe Thr Val Thr Pro Glu Phe Ile Phe Glu Gly Asp Asn Val Thr
275 280 285

Leu Glu Cys Glu Ser Glu Phe Val Ser Ser Asn Thr Ser Trp Phe Tyr
290 295 300

Gly Glu Lys Arg Ser Asp Ile Gln Asn Ser Asp Lys Phe Ser Ile His
305 310 315 320

Thr Ser Ile Ile Asn Asn Ile Ser Leu Val Thr Arg Leu Thr Ile Phe
325 330 335

Asn Phe Thr Gln His Asp Ala Gly Leu Tyr Gly Cys Asn Val Thr Leu
340 345 350

Asp Ile Phe Glu Tyr Gly Thr Val Arg Lys Leu Asp Val Thr Pro Ile
355 360 365

Arg Ile Leu Ala Lys Glu Glu Arg Lys Val Val Cys Asp Asn Asn Pro
370 375 380

Ile Ser Leu Asn Cys Cys Ser Glu Asn Ile Ala Asn Trp Ser Arg Ile
385 390 395 400

Glu Trp Lys Gln Glu Gly Lys Ile Asn Ile Glu Gly Thr Pro Glu Thr
405 410 415

Asp Leu Glu Ser Ser Cys Ser Thr Tyr Thr Leu Lys Ala Asp Gly Thr
420 425 430

Gln Cys Pro Ser Gly Ser Ser Gly Thr Thr Val Ile Tyr Thr Cys Glu
435 440 445

Phe Val Ser Val Tyr Gly Ala Lys Gly Ser Lys Asn Ile Ala Val Thr
450 455 460

Phe Thr Ser Val Ala Asn Leu Thr Ile Thr Pro Asp Pro Ile Ser Val
465 470 475 480

Ser Glu Gly Gln Ser Phe Ser Ile Thr Cys Leu Ser Asp Val Ser Ser
485 490 495

Phe Asp Glu Val Tyr Trp Asn Thr Ser Ala Gly Ile Lys Ile His Pro
500 505 510

Arg Phe Tyr Thr Met Arg Arg Tyr Arg Asp Gly Ala Glu Ser Val Leu
515 520 525

Thr Val Lys Thr Ser Thr Arg Glu Trp Asn Gly Thr Tyr His Cys Ile
530 535 540

Phe Arg Tyr Lys Asn Ser Tyr Ser Ile Ala Thr Lys Asp Val Thr Val
545 550 555 560

His Pro Leu Pro Leu Glu Ser Asp Ile Met Met Asp Pro Leu Glu Ala
565 570 575

Ser Gly Leu Cys Thr Ser Ser His Gln Phe Lys Cys Cys Ile Glu Glu
580 585 590

Asn Asp Gly Glu Glu Tyr Ile Val Thr Phe His Val Asp Ser Ser Ser
595 600 605

Phe Pro Ala Glu Arg Glu Val Ile Gly Lys Gln Ala Cys Tyr Thr Tyr
610 615 620

Ser Leu Pro Gly Lys Leu Pro Ser Arg Cys Pro Lys Asp Ile Asp Val
625 630 635 640

Phe Cys His Phe Thr Asn Ala Ala Asn Ser Ser Val Arg Ser Pro Ser
645 650 655

Met Lys Leu Thr Leu Val Pro Gly Lys Asn Ile Thr Cys Gln Asp Pro
660 665 670

Ile Ile Gly Ile Gly Glu Pro Gly Lys Val Ile Gln Lys Leu Cys Gln
675 680 685

Phe Ala Gly Val Ser Arg Ser Pro Gly Gln Thr Ile Gly Gly Thr Val
690 695 700

Thr Tyr Lys Cys Val Gly Ser Gln Trp Lys Glu Glu Thr Arg Ala Cys
705 710 715 720

Ile Ser Ala Pro Ile Asn Gly Leu Leu Gln Leu Ala Lys Ala Leu Ile
725 730 735

Lys Ser Pro Ser Gln Asp Gln Lys Leu Pro Lys Tyr Leu Arg Asp Leu
740 745 750

Ser Val Ser Thr Gly Lys Glu Glu Gln Asp Ile Arg Ser Ser Pro Gly
755 760 765

Ser Leu Gly Ala Ile Ile Ser Ile Leu Asp Leu Leu Ser Thr Val Pro
770 775 780

Thr Gln Val Asn Ser Glu Met Met Arg Asp Ile Leu Ala Thr Ile Asn
785 790 795 800

Val Ile Leu Asp Lys Ser Thr Leu Asn Ser Trp Glu Lys Leu Leu Gln
805 810 815

Gln Gln Ser Asn Gln Ser Ser Gln Phe Leu Gln Ser Val Glu Arg Phe
820 825 830

Ser Lys Ala Leu Glu Leu Gly Asp Ser Thr Pro Pro Phe Leu Phe His
835 840 845

Pro Asn Val Gln Met Lys Ser Met Val Ile Lys Arg Gly His Ala Gln
850 855 860

Met Tyr Gln Gln Lys Phe Val Phe Thr Asp Ser Asp Leu Trp Gly Asp
865 870 875 880

Val Ala Ile Asp Glu Cys Gln Leu Gly Ser Leu Gln Pro Asp Ser Ser
885 890 895

Ile Val Thr Val Ala Phe Pro Thr Leu Lys Ala Ile Leu Ala Gln Asp
900 905 910

Gly Gln Arg Lys Thr Pro Ser Asn Ser Leu Val Met Thr Thr Thr Val
915 920 925

Ser His Asn Ile Val Lys Pro Phe Arg Ile Ser Met Thr Phe Lys Asn
930 935 940

Asn His Arg Ser Gly Gly Lys Pro Gln Cys Val Phe Trp Asn Phe Ser
945 950 955 960

Leu Ala Asn Asn Thr Gly Gly Trp Asp Ser Ser Gly Cys Thr Val Glu
965 970 975

Asp Asp Gly Arg Asp Asn Arg Asp Arg Val Phe Cys Lys Cys Asn His
980 985 990

Leu Thr Ser Phe Ser Ile Leu Met Ser Pro Asp Ser Pro Asp Pro Gly
995 1000 1005

Ser Leu Leu Lys Ile Leu Leu Asp Ile Ile Ser Tyr Ile Gly Leu
1010 1015 1020

Gly Phe Ser Ile Val Ser Leu Ala Ala Cys Leu Val Val Glu Ala
1025 1030 1035

Met Val Trp Lys Ser Val Thr Lys Asn Arg Thr Ser Tyr Met Arg
1040 1045 1050

His Ile Cys Ile Val Asn Ile Ala Leu Cys Leu Leu Ile Ala Asp
1055 1060 1065

Ile Trp Phe Ile Val Ala Gly Ala Ile His Asp Gly His Tyr Pro
1070 1075 1080

Leu Asn Glu Thr Ala Cys Val Ala Ala Thr Phe Phe Ile His Phe
1085 1090 1095

Phe Tyr Leu Ser Val Phe Phe Trp Met Leu Thr Leu Gly Leu Met
1100 1105 1110

Leu Phe Tyr Arg Leu Ile Phe Ile Leu His Asp Ala Ser Lys Ser
1115 1120 1125

Thr Gln Lys Ala Ile Ala Phe Ser Leu Gly Tyr Gly Cys Pro Leu
1130 1135 1140

Ile Ile Ser Ser Ile Thr Val Gly Val Thr Gln Pro Gln Glu Val
1145 1150 1155

Tyr Met Arg Lys Asn Ala Cys Trp Leu Asn Trp Glu Asp Thr Arg
1160 1165 1170

Ala Leu Leu Ala Phe Ala Ile Pro Ala Leu Ile Ile Val Val Val
1175 1180 1185

Asn Val Ser Ile Thr Val Val Val Ile Thr Lys Ile Leu Arg Pro
1190 1195 1200

Ser Val Gly Asp Lys Pro Gly Lys Gln Glu Lys Ser Ser Leu Phe
1205 1210 1215

Gln Ile Ser Lys Ser Ile Gly Val Leu Thr Pro Leu Leu Gly Leu
1220 1225 1230

Thr Trp Gly Phe Gly Leu Ala Thr Val Ile Gln Gly Ser Asn Ala
1235 1240 1245

Val Phe His Ile Ile Phe Thr Leu Leu Asn Ala Phe Gln Gly Leu
1250 1255 1260

```
Phe Ile  Leu Leu Phe Gly Cys  Leu Trp Asp Gln Lys  Val Gln Glu
    1265                 1270                 1275

Ala Leu  Leu His Lys Phe Ser  Leu Ser Arg Trp Ser  Ser Gln His
    1280                 1285                 1290

Ser Lys  Ser Thr Ser Leu Gly  Ser Ser Thr Pro Val  Phe Ser Met
    1295                 1300                 1305

Ser Ser  Pro Ile Ser Arg Arg  Phe Asn Asn Leu Phe  Gly Lys Thr
    1310                 1315                 1320

Gly Thr  Tyr Asn Val Ser Thr  Pro Glu Thr Thr Ser  Ser Ser Val
    1325                 1330                 1335

Glu Asn  Ser Ser Ser Ala Tyr  Ser Leu Leu Asn
    1340                 1345
```

<210> 53
<211> 986
<212> PRT
<213> Homo sapiens

<400> 53

```
Cys Lys Lys Lys Ile Asp Val Met Pro Ile Gln Ile Leu Ala Asn Glu
1               5                   10                  15

Glu Met Lys Val Met Cys Asp Asn Asn Pro Val Ser Leu Asn Cys Cys
            20                  25                  30

Ser Gln Gly Asn Val Asn Trp Ser Lys Val Glu Trp Lys Gln Glu Gly
        35                  40                  45

Lys Ile Asn Ile Pro Gly Thr Pro Glu Thr Asp Ile Asp Ser Ser Cys
    50                  55                  60

Ser Arg Tyr Thr Leu Lys Ala Asp Gly Thr Gln Cys Pro Ser Gly Ser
65                  70                  75                  80

Ser Gly Thr Thr Val Ile Tyr Thr Cys Glu Phe Ile Ser Ala Tyr Gly
                85                  90                  95

Ala Arg Gly Ser Ala Asn Ile Lys Val Thr Phe Ile Ser Val Ala Asn
            100                 105                 110

Leu Thr Ile Thr Pro Asp Pro Ile Ser Val Ser Glu Gly Gln Asn Phe
        115                 120                 125

Ser Ile Lys Cys Ile Ser Asp Val Ser Asn Tyr Asp Glu Val Tyr Trp
    130                 135                 140

Asn Thr Ser Ala Gly Ile Lys Ile Tyr Gln Arg Phe Tyr Thr Thr Arg
145                 150                 155                 160

Arg Tyr Leu Asp Gly Ala Glu Ser Val Leu Thr Val Lys Thr Ser Thr
                165                 170                 175

Arg Glu Trp Asn Gly Thr Tyr His Cys Ile Phe Arg Tyr Lys Asn Ser
            180                 185                 190

Tyr Ser Ile Ala Thr Lys Asp Val Ile Val His Pro Leu Pro Leu Lys
        195                 200                 205
```

Leu Asn Ile Met Val Asp Pro Leu Glu Ala Thr Val Ser Cys Ser Gly
210 215 220

Ser His His Ile Lys Cys Cys Ile Glu Glu Asp Gly Asp Tyr Lys Val
225 230 235 240

Thr Phe His Met Gly Ser Ser Ser Leu Pro Ala Ala Lys Glu Val Asn
245 250 255

Lys Lys Gln Val Cys Tyr Lys His Asn Phe Asn Ala Ser Ser Val Ser
260 265 270

Trp Cys Ser Lys Thr Val Asp Val Cys Cys His Phe Thr Asn Ala Ala
275 280 285

Asn Asn Ser Val Trp Ser Pro Ser Met Lys Leu Asn Leu Val Pro Gly
290 295 300

Glu Asn Ile Thr Cys Gln Asp Pro Val Ile Gly Val Gly Glu Pro Gly
305 310 315 320

Lys Val Ile Gln Lys Leu Cys Arg Phe Ser Asn Val Pro Ser Ser Pro
325 330 335

Glu Ser Pro Ile Gly Gly Thr Ile Thr Tyr Lys Cys Val Gly Ser Gln
340 345 350

Trp Glu Glu Lys Arg Asn Asp Cys Ile Ser Ala Pro Ile Asn Ser Leu
355 360 365

Leu Gln Met Ala Lys Ala Leu Ile Lys Ser Pro Ser Gln Asp Glu Met
370 375 380

Leu Pro Thr Tyr Leu Lys Asp Leu Ser Ile Ser Ile Asp Lys Ala Glu
385 390 395 400

His Glu Ile Ser Ser Ser Pro Gly Ser Leu Gly Ala Ile Ile Asn Ile
405 410 415

Leu Asp Leu Leu Ser Thr Val Pro Thr Gln Val Asn Ser Glu Met Met
420 425 430

Thr His Val Leu Ser Thr Val Asn Val Ile Leu Gly Lys Pro Val Leu
435 440 445

Asn Thr Trp Lys Val Leu Gln Gln Gln Trp Thr Asn Gln Ser Ser Gln
450 455 460

Leu Leu His Ser Val Glu Arg Phe Ser Gln Ala Leu Gln Ser Gly Asp
465 470 475 480

Ser Pro Pro Leu Ser Phe Ser Gln Thr Asn Val Gln Met Ser Ser Thr
485 490 495

Val Ile Lys Ser Ser His Pro Glu Thr Tyr Gln Gln Arg Phe Val Phe
500 505 510

Pro Tyr Phe Asp Leu Trp Gly Asn Val Val Ile Asp Lys Ser Tyr Leu
515 520 525

Glu Asn Leu Gln Ser Asp Ser Ser Ile Val Thr Met Ala Phe Pro Thr
530 535 540

Leu Gln Ala Ile Leu Ala Gln Asp Ile Gln Glu Asn Asn Phe Ala Glu
545 550 555 560

Ser Leu Val Met Thr Thr Thr Val Ser His Asn Thr Thr Met Pro Phe
565 570 575

Arg Ile Ser Met Thr Phe Lys Asn Asn Ser Pro Ser Gly Gly Glu Thr
580 585 590

Lys Cys Val Phe Trp Asn Phe Arg Leu Ala Asn Asn Thr Gly Gly Trp
595 600 605

Asp Ser Ser Gly Cys Tyr Val Glu Glu Gly Asp Gly Asp Asn Val Thr
610 615 620

Cys Ile Cys Asp His Leu Thr Ser Phe Ser Ile Leu Met Ser Pro Asp
625 630 635 640

Ser Pro Asp Pro Ser Ser Leu Leu Gly Ile Leu Leu Asp Ile Ile Ser
645 650 655

Tyr Val Gly Val Gly Phe Ser Ile Leu Ser Leu Ala Ala Cys Leu Val
660 665 670

Val Glu Ala Val Val Trp Lys Ser Val Thr Lys Asn Arg Thr Ser Tyr
675 680 685

Met Arg His Thr Cys Ile Val Asn Ile Ala Ala Ser Leu Leu Val Ala
690 695 700

Asn Thr Trp Phe Ile Val Val Ala Ala Ile Gln Asp Asn Arg Tyr Ile
705 710 715 720

Leu Cys Lys Thr Ala Cys Val Ala Ala Thr Phe Phe Ile His Phe Phe
725 730 735

Tyr Leu Ser Val Phe Phe Trp Met Leu Thr Leu Gly Leu Met Leu Phe
740 745 750

Tyr Arg Leu Val Phe Ile Leu His Glu Thr Ser Arg Ser Thr Gln Lys
755 760 765

Ala Ile Ala Phe Cys Leu Gly Tyr Gly Cys Pro Leu Ala Ile Ser Val
770 775 780

Ile Thr Leu Gly Ala Thr Gln Pro Arg Glu Val Tyr Thr Arg Lys Asn
785 790 795 800

Val Cys Trp Leu Asn Trp Glu Asp Thr Lys Ala Leu Leu Ala Phe Ala
805 810 815

Ile Pro Ala Leu Ile Ile Val Val Val Asn Ile Thr Ile Thr Ile Val
820 825 830

Val Ile Thr Lys Ile Leu Arg Pro Ser Ile Gly Asp Lys Pro Cys Lys
835 840 845

Gln Glu Lys Ser Ser Leu Phe Gln Ile Ser Lys Ser Ile Gly Val Leu
850 855 860

Thr Pro Leu Leu Gly Leu Thr Trp Gly Phe Gly Leu Thr Thr Val Phe
865 870 875 880

Pro Gly Thr Asn Leu Val Phe His Ile Ile Phe Ala Ile Leu Asn Val
885 890 895

Phe Gln Gly Leu Phe Ile Leu Leu Phe Gly Cys Leu Trp Asp Leu Lys
900 905 910

Val Gln Glu Ala Leu Leu Asn Lys Phe Ser Leu Ser Arg Trp Ser Ser
915 920 925

Gln His Ser Lys Ser Thr Ser Leu Gly Ser Ser Thr Pro Val Phe Ser
930 935 940

Met Ser Ser Pro Ile Ser Arg Arg Phe Asn Asn Leu Phe Gly Lys Thr
945 950 955 960

Gly Thr Tyr Asn Val Ser Thr Pro Glu Ala Thr Ser Ser Ser Leu Glu
965 970 975

Asn Ser Ser Ser Ala Ser Ser Leu Leu Asn
980 985

<210> 54
<211> 322
<212> PRT
<213> Mus musculus

<400> 54

Leu Ser Pro Ala Val Pro Pro Tyr Val Lys Leu Gly Leu Thr Ala Val
1 5 10 15

Tyr Thr Val Phe Tyr Ala Leu Leu Phe Val Phe Ile Tyr Ala Gln Leu
20 25 30

Trp Leu Val Leu Arg Tyr Arg His Lys Arg Leu Ser Tyr Gln Ser Val
35 40 45

Phe Leu Phe Leu Cys Leu Phe Trp Ala Ser Leu Arg Thr Val Leu Phe
50 55 60

Ser Phe Tyr Phe Arg Asp Phe Val Ala Ala Asn Ser Phe Ser Pro Phe
65 70 75 80

Val Phe Trp Leu Leu Tyr Cys Phe Pro Val Cys Leu Gln Phe Phe Thr
85 90 95

Leu Thr Leu Met Asn Leu Tyr Phe Thr Gln Val Ile Phe Lys Ala Lys
100 105 110

Ser Lys Tyr Ser Pro Glu Leu Leu Lys Tyr Arg Leu Pro Leu Tyr Leu
115 120 125

Ala Ser Leu Phe Ile Ser Leu Val Phe Leu Leu Val Asn Leu Thr Cys
130 135 140

Ala Val Leu Val Lys Thr Gly Asp Trp Asp Arg Lys Val Ile Val Ser
145 150 155 160

Val Arg Val Ala Ile Asn Asp Thr Leu Phe Val Leu Cys Ala Ile Ser
165 170 175

Leu Ser Ile Cys Leu Tyr Lys Ile Ser Lys Met Ser Leu Ala Asn Ile

180 185 190

Tyr Leu Glu Ser Lys Gly Ser Ser Val Cys Gln Val Thr Ala Ile Gly
195 200 205

Val Thr Val Ile Leu Leu Tyr Ala Ser Arg Ala Cys Tyr Asn Leu Phe
210 215 220

Ile Leu Ser Phe Ser Gln Ile Lys Asn Val His Ser Phe Asp Tyr Asp
225 230 235 240

Trp Tyr Asn Val Ser Asp Gln Ala Asp Leu Lys Ser Gln Leu Gly Asp
245 250 255

Ala Gly Tyr Val Val Phe Gly Val Val Leu Phe Val Trp Glu Leu Leu
260 265 270

Pro Thr Thr Leu Val Val Tyr Phe Phe Arg Val Arg Asn Pro Thr Lys
275 280 285

Asp Leu Thr Asn Pro Gly Met Val Pro Ser His Gly Phe Ser Pro Arg
290 295 300

Ser Tyr Phe Phe Asp Asn Pro Arg Arg Tyr Asp Ser Asp Asp Asp Leu
305 310 315 320

Ala Trp

<210> 55
<211> 392
<212> PRT
<213> Homo sapiens

<400> 55

Met Arg Ser Thr Thr Leu Leu Ala Leu Leu Ala Leu Val Leu Leu Tyr
1 5 10 15

Leu Val Ser Gly Ala Leu Val Phe Arg Ala Leu Glu Gln Pro His Glu
20 25 30

Gln Gln Ala Gln Arg Glu Leu Gly Glu Val Arg Glu Lys Phe Leu Arg
35 40 45

Ala His Pro Cys Val Ser Asp Gln Glu Leu Gly Leu Leu Ile Lys Glu
50 55 60

Val Ala Asp Ala Leu Gly Gly Gly Ala Asp Pro Glu Thr Asn Ser Thr
65 70 75 80

Ser Asn Ser Ser His Ser Ala Trp Asp Leu Gly Ser Ala Phe Phe Phe
85 90 95

Ser Gly Thr Ile Ile Thr Thr Ile Gly Tyr Gly Asn Val Ala Leu Arg
100 105 110

Thr Asp Ala Gly Arg Leu Phe Cys Ile Phe Tyr Ala Leu Val Gly Ile
115 120 125

Pro Leu Phe Gly Ile Leu Leu Ala Gly Val Gly Asp Arg Leu Gly Ser
130 135 140

Ser Leu Arg His Gly Ile Gly His Ile Glu Ala Ile Phe Leu Lys Trp
145 150 155 160

His Val Pro Pro Glu Leu Val Arg Val Leu Ser Ala Met Leu Phe Leu
165 170 175

Leu Ile Gly Cys Leu Leu Phe Val Leu Thr Pro Thr Phe Val Phe Cys
180 185 190

Tyr Met Glu Asp Trp Ser Lys Leu Glu Ala Ile Tyr Phe Val Ile Val
195 200 205

Thr Leu Thr Thr Val Gly Phe Gly Asp Tyr Val Ala Gly Ala Asp Pro
210 215 220

Arg Gln Asp Ser Pro Ala Tyr Gln Pro Leu Val Trp Phe Trp Ile Leu
225 230 235 240

Leu Gly Leu Ala Tyr Phe Ala Ser Val Leu Thr Thr Ile Gly Asn Trp
245 250 255

Leu Arg Val Val Ser Arg Arg Thr Arg Ala Glu Met Gly Gly Leu Thr
260 265 270

Ala Gln Ala Ala Ser Trp Thr Gly Thr Val Thr Ala Arg Val Thr Gln
275 280 285

Arg Ala Gly Pro Ala Ala Pro Pro Pro Glu Lys Glu Gln Pro Leu Leu
290 295 300

Pro Pro Pro Pro Cys Pro Ala Gln Pro Leu Gly Arg Pro Arg Ser Pro
305 310 315 320

Ser Pro Pro Glu Lys Ala Gln Pro Pro Ser Pro Pro Thr Ala Ser Ala
325 330 335

Leu Asp Tyr Pro Ser Glu Asn Leu Ala Phe Ile Asp Glu Ser Ser Asp
340 345 350

Thr Gln Ser Glu Arg Gly Cys Pro Leu Pro Arg Ala Pro Arg Gly Arg
355 360 365

Arg Arg Pro Asn Pro Pro Arg Lys Pro Val Arg Pro Arg Gly Pro Gly
370 375 380

Arg Pro Arg Asp Lys Gly Val Pro
385 390

<210> 56
<211> 166
<212> PRT
<213> Homo sapiens

<400> 56

Met Ser Tyr Asp Arg Tyr Met Ala Ile Cys His Pro Leu Gln Tyr Ser
1 5 10 15

Val Ile Met Arg Trp Gly Val Cys Thr Val Leu Ala Val Thr Ser Trp
20 25 30

Ala Cys Gly Ser Leu Leu Ala Leu Val His Val Val Leu Ile Leu Arg
35 40 45

Leu Pro Phe Cys Gly Pro His Glu Ile Asn His Phe Phe Cys Glu Ile
50 55 60

Leu Ser Val Leu Lys Leu Ala Cys Ala Asp Thr Trp Leu Asn Gln Val
65 70 75 80

Val Ile Phe Ala Ala Ser Val Phe Ile Leu Val Gly Pro Leu Cys Leu
85 90 95

Val Leu Val Ser Tyr Ser Arg Ile Leu Ala Ala Ile Leu Arg Ile Gln
100 105 110

Ser Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser Ser His Leu
115 120 125

Cys Met Val Gly Leu Phe Phe Gly Ser Ala Ile Val Met Tyr Met Ala
130 135 140

Pro Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val Leu Ser Leu Phe
145 150 155 160

Tyr Ser Leu Phe Asn Pro
165

<210> 57
<211> 171
<212> PRT
<213> Homo sapiens

<400> 57

Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr Phe
1 5 10 15

Ile Ile Met Thr Trp Lys Val Cys Ile Thr Leu Gly Ile Thr Ser Trp
20 25 30

Thr Cys Gly Ser Leu Leu Ala Met Val His Val Ser Leu Ile Leu Arg
35 40 45

Leu Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe Phe Cys Glu Ile
50 55 60

Leu Ser Val Leu Arg Leu Ala Cys Ala Asp Thr Trp Leu Asn Gln Val
65 70 75 80

Val Ile Phe Glu Ala Cys Met Phe Ile Leu Val Gly Pro Leu Cys Leu
85 90 95

Val Leu Val Ser Tyr Ser His Ile Leu Gly Gly Ile Leu Arg Ile Gln
100 105 110

Ser Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser Ser His Leu
115 120 125

Cys Val Val Gly Leu Phe Phe Gly Ser Ala Ile Val Met Tyr Met Ala
130 135 140

Pro Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val Leu Phe Leu Ile
145 150 155 160

Leu Gln Phe Leu Ser Thr Pro Met Leu Lys Pro

165 170

<210> 58
<211> 304
<212> PRT
<213> Homo sapiens

<400> 58

Met Gly Asp Asn Ile Thr Ser Ile Arg Glu Phe Leu Leu Leu Gly Phe
1 5 10 15

Pro Val Gly Pro Arg Ile Gln Met Leu Leu Phe Gly Leu Phe Ser Leu
20 25 30

Phe Tyr Val Phe Thr Leu Leu Gly Asn Gly Thr Ile Leu Gly Leu Ile
35 40 45

Ser Leu Asp Ser Arg Leu His Ala Pro Met Tyr Phe Phe Leu Ser His
50 55 60

Leu Ala Val Val Asp Ile Ala Tyr Ala Cys Asn Thr Val Pro Arg Met
65 70 75 80

Leu Val Asn Leu Leu His Pro Ala Lys Pro Ile Ser Phe Ala Gly Arg
85 90 95

Met Met Gln Thr Phe Leu Phe Ser Thr Phe Ala Val Thr Glu Cys Leu
100 105 110

Leu Leu Val Val Met Ser Tyr Asp Leu Tyr Val Ala Ile Cys His Pro
115 120 125

Leu Arg Tyr Leu Ala Ile Met Thr Trp Arg Val Cys Ile Thr Leu Ala
130 135 140

Val Thr Ser Trp Thr Thr Gly Val Leu Leu Ser Leu Ile His Leu Val
145 150 155 160

Leu Leu Leu Pro Leu Pro Phe Cys Arg Pro Gln Lys Ile Tyr His Phe
165 170 175

Phe Cys Glu Ile Leu Ala Val Leu Lys Leu Ala Cys Ala Asp Thr His
180 185 190

Ile Asn Glu Asn Met Val Leu Ala Gly Ala Ile Ser Gly Leu Val Gly
195 200 205

Pro Leu Ser Thr Ile Val Val Ser Tyr Met Cys Ile Leu Cys Ala Ile
210 215 220

Leu Gln Ile Gln Ser Arg Glu Val Gln Arg Lys Ala Phe Arg Thr Cys
225 230 235 240

Phe Ser His Leu Cys Val Ile Gly Leu Val Tyr Gly Thr Ala Ile Ile
245 250 255

Met Tyr Val Gly Pro Arg Tyr Gly Asn Pro Lys Glu Gln Lys Lys Tyr
260 265 270

Leu Leu Leu Phe His Ser Leu Phe Asn Pro Met Leu Asn Pro Leu Ile
275 280 285

Cys Ser Leu Arg Asn Ser Glu Val Lys Asn Thr Leu Lys Arg Val Leu
290 295 300

<210> 59
<211> 287
<212> PRT
<213> Homo sapiens

<400> 59

Met Val Lys Asn Gln Thr Met Val Thr Glu Phe Leu Leu Leu Gly Phe
1 5 10 15

Leu Leu Gly Pro Arg Ile Gln Met Leu Leu Phe Gly Leu Phe Ser Leu
20 25 30

Phe Tyr Val Phe Thr Leu Leu Gly Asn Gly Thr Ile Leu Gly Leu Ile
35 40 45

Ser Leu Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ser His
50 55 60

Leu Ala Val Val Asn Ile Ala Tyr Ala Cys Asn Thr Val Pro Gln Met
65 70 75 80

Leu Val Asn Leu Leu His Pro Ala Lys Pro Ile Ser Phe Ala Gly Cys
85 90 95

Met Thr Leu Asp Phe Leu Phe Leu Ser Phe Ala His Thr Glu Cys Leu
100 105 110

Leu Leu Val Leu Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro
115 120 125

Leu Arg Tyr Phe Ile Ile Met Thr Trp Lys Val Cys Ile Thr Leu Gly
130 135 140

Ile Thr Ser Trp Thr Cys Gly Ser Leu Leu Ala Met Val His Val Ser
145 150 155 160

Leu Ile Leu Arg Leu Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe
165 170 175

Phe Cys Glu Ile Leu Ser Val Leu Arg Leu Ala Cys Ala Asp Thr Trp
180 185 190

Leu Asn Gln Val Val Ile Phe Glu Ala Cys Met Phe Ile Leu Val Gly
195 200 205

Pro Leu Cys Leu Val Leu Val Ser Tyr Ser His Ile Leu Gly Gly Ile
210 215 220

Leu Arg Ile Gln Ser Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys
225 230 235 240

Ser Ser His Leu Cys Val Val Gly Leu Phe Phe Gly Ser Ala Ile Val
245 250 255

Met Tyr Met Ala Pro Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val
260 265 270

Leu Phe Leu Ile Leu Gln Phe Leu Ser Thr Pro Met Leu Lys Pro
275 280 285

<210> 60
<211> 5714
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (272)..(4312)

<400> 60
```
ctgcacgacc ggtccggaat tcccgggtcg acgatttcgt gatcatagct gggggaggct      60

gagcgtggga gcggtgctgc cagtcctgcc tgaaaacgcg aaatgagtct tgcttggttc     120

tccctccact gggcgtgaga gcccctgccc aggaggccca ggacaaatgg ccccatagtg     180

gaaactggga agcttttagg catctgatca gagcgggagc cagccggggg accacagtgc     240

tggacaggcc aaccaactca aacttgaaga c atg aaa tcc cca agg aga acc        292
                                   Met Lys Ser Pro Arg Arg Thr
                                   1               5

act ttg tgc ctc atg ttt att gtg att tat tct tcc aaa gct gca ctg      340
Thr Leu Cys Leu Met Phe Ile Val Ile Tyr Ser Ser Lys Ala Ala Leu
        10                  15                  20

aac tgg aat tac gag tct act att cat cct ttg agt ctt cat gaa cat      388
Asn Trp Asn Tyr Glu Ser Thr Ile His Pro Leu Ser Leu His Glu His
    25                  30                  35

gaa cca gct ggt gaa gag gca ctg agg caa aaa cga gcc gtt gcc aca      436
Glu Pro Ala Gly Glu Glu Ala Leu Arg Gln Lys Arg Ala Val Ala Thr
40                  45                  50                  55

aaa agt cct acg gct gaa gaa tac act gtt aat att gag atc agt ttt      484
Lys Ser Pro Thr Ala Glu Glu Tyr Thr Val Asn Ile Glu Ile Ser Phe
                60                  65                  70

gaa aat gca tcc ttc ctg gat cct atc aaa gcc tac ttg aac agc ctc      532
Glu Asn Ala Ser Phe Leu Asp Pro Ile Lys Ala Tyr Leu Asn Ser Leu
            75                  80                  85

agt ttt cca att cat ggg aat aac act gac caa att act gac att ttg      580
Ser Phe Pro Ile His Gly Asn Asn Thr Asp Gln Ile Thr Asp Ile Leu
        90                  95                  100

agc ata aat gtg aca aca gtc tgc aga cct gct gga aat gaa atc tgg      628
Ser Ile Asn Val Thr Thr Val Cys Arg Pro Ala Gly Asn Glu Ile Trp
    105                 110                 115

tgc tcc tgc gag aca ggt tat ggg tgg cct cgg gaa agg tgt ctt cac      676
Cys Ser Cys Glu Thr Gly Tyr Gly Trp Pro Arg Glu Arg Cys Leu His
120                 125                 130                 135

aat ctc att tgt caa gag cgt gac gtc ttc ctc cca ggg cac cat tgc      724
Asn Leu Ile Cys Gln Glu Arg Asp Val Phe Leu Pro Gly His His Cys
                140                 145                 150

agt tgc ctt aaa gaa ctg cct ccc aat gga cct ttt tgc ctg ctt cag      772
Ser Cys Leu Lys Glu Leu Pro Pro Asn Gly Pro Phe Cys Leu Leu Gln
            155                 160                 165
```

gaa gat gtt acc ctg aac atg aga gtc aga cta aat gta ggc ttt caa 820
Glu Asp Val Thr Leu Asn Met Arg Val Arg Leu Asn Val Gly Phe Gln
170 175 180

gaa gac ctc atg aac act tcc tcc gcc ctc tat agg tcc tac aag acc 868
Glu Asp Leu Met Asn Thr Ser Ser Ala Leu Tyr Arg Ser Tyr Lys Thr
185 190 195

gac ttg gaa aca gcg ttc cgg aag ggt tac gga att tta cca ggc ttc 916
Asp Leu Glu Thr Ala Phe Arg Lys Gly Tyr Gly Ile Leu Pro Gly Phe
200 205 210 215

aag ggc gtg act gtg aca ggg ttc aag tct gga agt gtg gtt gtg aca 964
Lys Gly Val Thr Val Thr Gly Phe Lys Ser Gly Ser Val Val Val Thr
220 225 230

tat gaa gtc aag act aca cca cca tca ctt gag tta ata cat aaa gcc 1012
Tyr Glu Val Lys Thr Thr Pro Pro Ser Leu Glu Leu Ile His Lys Ala
235 240 245

aat gaa caa gtt gta cag agc ctc aat cag acc tac aaa atg gac tac 1060
Asn Glu Gln Val Val Gln Ser Leu Asn Gln Thr Tyr Lys Met Asp Tyr
250 255 260

aac tcc ttt caa gca gtt act atc aat gaa agc aat ttc ttt gtc aca 1108
Asn Ser Phe Gln Ala Val Thr Ile Asn Glu Ser Asn Phe Phe Val Thr
265 270 275

cca gaa atc atc ttt gaa ggg gac aca gtc agt ctg gtg tgt gaa aag 1156
Pro Glu Ile Ile Phe Glu Gly Asp Thr Val Ser Leu Val Cys Glu Lys
280 285 290 295

gaa gtt ttg tcc tcc aat gtg tct tgg cgc tat gaa gaa cag cag ttg 1204
Glu Val Leu Ser Ser Asn Val Ser Trp Arg Tyr Glu Glu Gln Gln Leu
300 305 310

gaa atc cag aac agc agc aga ttc tcg att tac acc gca ctt ttc aac 1252
Glu Ile Gln Asn Ser Ser Arg Phe Ser Ile Tyr Thr Ala Leu Phe Asn
315 320 325

aac atg act tcg gtg tcc aag ctc acc atc cac aac atc act cca ggt 1300
Asn Met Thr Ser Val Ser Lys Leu Thr Ile His Asn Ile Thr Pro Gly
330 335 340

gat gca ggt gaa tat gtt tgc aaa ctg ata tta gac att ttt gaa tat 1348
Asp Ala Gly Glu Tyr Val Cys Lys Leu Ile Leu Asp Ile Phe Glu Tyr
345 350 355

gag tgc aag aag aaa ata gat gtt atg ccc atc caa att ttg gca aat 1396
Glu Cys Lys Lys Lys Ile Asp Val Met Pro Ile Gln Ile Leu Ala Asn
360 365 370 375

gaa gaa atg aag gtg atg tgc gac aac aat cct gta tct ttg aac tgc 1444
Glu Glu Met Lys Val Met Cys Asp Asn Asn Pro Val Ser Leu Asn Cys
380 385 390

tgc agt cag ggt aat gtt aat tgg agc aaa gta gaa tgg aag cag gaa 1492
Cys Ser Gln Gly Asn Val Asn Trp Ser Lys Val Glu Trp Lys Gln Glu
395 400 405

gga aaa ata aat att cca gga acc cct gag aca gac ata gat tct agc 1540
Gly Lys Ile Asn Ile Pro Gly Thr Pro Glu Thr Asp Ile Asp Ser Ser
410 415 420

```
tgc agc aga tac acc ctc aag gct gat gga acc cag tgc cca agc ggg      1588
Cys Ser Arg Tyr Thr Leu Lys Ala Asp Gly Thr Gln Cys Pro Ser Gly
    425                 430                 435

tcg tct gga aca aca gtc atc tac act tgt gag ttc atc agt gcc tat      1636
Ser Ser Gly Thr Thr Val Ile Tyr Thr Cys Glu Phe Ile Ser Ala Tyr
440                 445                 450                 455

gga gcc aga ggc agt gca aac ata aaa gtg aca ttc atc tct gtg gcc      1684
Gly Ala Arg Gly Ser Ala Asn Ile Lys Val Thr Phe Ile Ser Val Ala
                460                 465                 470

aat cta aca ata acc ccg gac cca att tct gtt tct gag gga caa aac      1732
Asn Leu Thr Ile Thr Pro Asp Pro Ile Ser Val Ser Glu Gly Gln Asn
            475                 480                 485

ttt tct ata aaa tgc atc agt gat gtg agt aac tat gat gag gtt tat      1780
Phe Ser Ile Lys Cys Ile Ser Asp Val Ser Asn Tyr Asp Glu Val Tyr
        490                 495                 500

tgg aac act tct gct gga att aaa ata tac caa aga ttt tat acc acg      1828
Trp Asn Thr Ser Ala Gly Ile Lys Ile Tyr Gln Arg Phe Tyr Thr Thr
    505                 510                 515

agg agg tat ctt gat gga gca gaa tca gta ctg aca gtc aag acc tcg      1876
Arg Arg Tyr Leu Asp Gly Ala Glu Ser Val Leu Thr Val Lys Thr Ser
520                 525                 530                 535

acc agg gag tgg aat gga acc tat cac tgc ata ttt aga tat aag aat      1924
Thr Arg Glu Trp Asn Gly Thr Tyr His Cys Ile Phe Arg Tyr Lys Asn
                540                 545                 550

tca tac agt att gca acc aaa gac gtc att gtt cac ccg ctg cct cta      1972
Ser Tyr Ser Ile Ala Thr Lys Asp Val Ile Val His Pro Leu Pro Leu
            555                 560                 565

aag ctg aac atc atg gtt gat cct ttg gaa gct act gtt tca tgc agt      2020
Lys Leu Asn Ile Met Val Asp Pro Leu Glu Ala Thr Val Ser Cys Ser
        570                 575                 580

ggt tcc cat cac atc aag tgc tgc ata gag gag gat gga gac tac aaa      2068
Gly Ser His His Ile Lys Cys Cys Ile Glu Glu Asp Gly Asp Tyr Lys
    585                 590                 595

gtt act ttc cat atg ggt tcc tca tcc ctt cct gct gca aaa gaa gtt      2116
Val Thr Phe His Met Gly Ser Ser Ser Leu Pro Ala Ala Lys Glu Val
600                 605                 610                 615

aac aaa aaa caa gtg tgc tac aaa cac aat ttc aat gca agc tca gtt      2164
Asn Lys Lys Gln Val Cys Tyr Lys His Asn Phe Asn Ala Ser Ser Val
                620                 625                 630

tcc tgg tgt tca aaa act gtt gat gtg tgt tgt cac ttt acc aat gct      2212
Ser Trp Cys Ser Lys Thr Val Asp Val Cys Cys His Phe Thr Asn Ala
            635                 640                 645

gct aat aat tca gtt tgg agc cca tct atg aag ctg aat ctg gtt cct      2260
Ala Asn Asn Ser Val Trp Ser Pro Ser Met Lys Leu Asn Leu Val Pro
        650                 655                 660

ggg gaa aac atc aca tgc cag gat ccc gta ata ggt gtc gga gag ccg      2308
Gly Glu Asn Ile Thr Cys Gln Asp Pro Val Ile Gly Val Gly Glu Pro
```

665 670 675

ggg aaa gtc atc cag aag cta tgc cgg ttc tca aac gtt ccc agc agc 2356
Gly Lys Val Ile Gln Lys Leu Cys Arg Phe Ser Asn Val Pro Ser Ser
680 685 690 695

cct gag agt ccc att ggc ggg acc atc act tac aaa tgt gta ggc tcc 2404
Pro Glu Ser Pro Ile Gly Gly Thr Ile Thr Tyr Lys Cys Val Gly Ser
700 705 710

cag tgg gag gag aag aga aat gac tgc atc tct gcc cca ata aac agt 2452
Gln Trp Glu Glu Lys Arg Asn Asp Cys Ile Ser Ala Pro Ile Asn Ser
715 720 725

ctg ctc cag atg gct aag gct ttg atc aag agc ccc tct cag gat gag 2500
Leu Leu Gln Met Ala Lys Ala Leu Ile Lys Ser Pro Ser Gln Asp Glu
730 735 740

atg ctc cct aca tac ctg aag gat ctt tct att agc ata gac aaa gcg 2548
Met Leu Pro Thr Tyr Leu Lys Asp Leu Ser Ile Ser Ile Asp Lys Ala
745 750 755

gaa cat gaa atc agc tct tct cct ggg agt ctg gga gcc att att aac 2596
Glu His Glu Ile Ser Ser Ser Pro Gly Ser Leu Gly Ala Ile Ile Asn
760 765 770 775

atc ctt gat ctg ctc tca aca gtt cca acc caa gta aat tca gaa atg 2644
Ile Leu Asp Leu Leu Ser Thr Val Pro Thr Gln Val Asn Ser Glu Met
780 785 790

atg acg cac gtg ctc tct acg gtt aat gtc atc ctt ggc aag ccc gtc 2692
Met Thr His Val Leu Ser Thr Val Asn Val Ile Leu Gly Lys Pro Val
795 800 805

ttg aac acc tgg aag gtt tta caa cag caa tgg acc aat cag agt tca 2740
Leu Asn Thr Trp Lys Val Leu Gln Gln Gln Trp Thr Asn Gln Ser Ser
810 815 820

cag cta cta cat tca gtg gaa aga ttt tcc caa gca tta cag tca gga 2788
Gln Leu Leu His Ser Val Glu Arg Phe Ser Gln Ala Leu Gln Ser Gly
825 830 835

gat agc cct cct ttg tcc ttc tcc caa act aat gtg cag atg agc agc 2836
Asp Ser Pro Pro Leu Ser Phe Ser Gln Thr Asn Val Gln Met Ser Ser
840 845 850 855

acg gta atc aag tcc agc cac cca gaa acc tat caa cag agg ttt gtt 2884
Thr Val Ile Lys Ser Ser His Pro Glu Thr Tyr Gln Gln Arg Phe Val
860 865 870

ttc cca tac ttt gac ctc tgg ggc aat gtg gtc att gac aag agc tac 2932
Phe Pro Tyr Phe Asp Leu Trp Gly Asn Val Val Ile Asp Lys Ser Tyr
875 880 885

cta gaa aac ttg cag tcg gat tcg tct att gtc acc atg gct ttc cca 2980
Leu Glu Asn Leu Gln Ser Asp Ser Ser Ile Val Thr Met Ala Phe Pro
890 895 900

act ctc caa gcc atc ctt gct cag gat atc cag gaa aat aac ttt gca 3028
Thr Leu Gln Ala Ile Leu Ala Gln Asp Ile Gln Glu Asn Asn Phe Ala
905 910 915

gag agc tta gtg atg aca acc act gtc agc cac aat acg act atg cca 3076

Glu Ser Leu Val Met Thr Thr Thr Val Ser His Asn Thr Thr Met Pro
920 925 930 935

ttc agg att tca atg act ttt aag aac aat agc cct tca ggc ggc gaa 3124
Phe Arg Ile Ser Met Thr Phe Lys Asn Asn Ser Pro Ser Gly Gly Glu
940 945 950

acg aag tgt gtc ttc tgg aac ttc agg ctt gcc aac aac aca ggg ggg 3172
Thr Lys Cys Val Phe Trp Asn Phe Arg Leu Ala Asn Asn Thr Gly Gly
955 960 965

tgg gac agc agt ggg tgc tat gtt gaa gaa ggt gat ggg gac aat gtc 3220
Trp Asp Ser Ser Gly Cys Tyr Val Glu Glu Gly Asp Gly Asp Asn Val
970 975 980

acc tgt atc tgt gac cac cta aca tca ttc tcc atc ctc atg tcc cct 3268
Thr Cys Ile Cys Asp His Leu Thr Ser Phe Ser Ile Leu Met Ser Pro
985 990 995

gac tcc cca gat cct agt tct ctc ctg gga ata ctc ctg gat att 3313
Asp Ser Pro Asp Pro Ser Ser Leu Leu Gly Ile Leu Leu Asp Ile
1000 1005 1010

att tct tat gtt ggg gtg ggc ttt tcc atc ttg agc ttg gca gcc 3358
Ile Ser Tyr Val Gly Val Gly Phe Ser Ile Leu Ser Leu Ala Ala
1015 1020 1025

tgt cta gtt gtg gaa gct gtg gtg tgg aaa tcg gtg acc aag aat 3403
Cys Leu Val Val Glu Ala Val Val Trp Lys Ser Val Thr Lys Asn
1030 1035 1040

cgg act tct tat atg cgc cac acc tgc ata gtg aat atc gct gcc 3448
Arg Thr Ser Tyr Met Arg His Thr Cys Ile Val Asn Ile Ala Ala
1045 1050 1055

tcc ctt ctg gtc gcc aac acc tgg ttc att gtg gtc gct gcc atc 3493
Ser Leu Leu Val Ala Asn Thr Trp Phe Ile Val Val Ala Ala Ile
1060 1065 1070

cag gac aat cgc tac ata ctc tgc aag aca gcc tgt gtg gct gcc 3538
Gln Asp Asn Arg Tyr Ile Leu Cys Lys Thr Ala Cys Val Ala Ala
1075 1080 1085

acc ttc ttc atc cac ttc ttc tac ctc agc gtc ttc ttc tgg atg 3583
Thr Phe Phe Ile His Phe Phe Tyr Leu Ser Val Phe Phe Trp Met
1090 1095 1100

ctg aca ctg ggc ctc atg ctg ttc tat cgc ctg gtt ttc att ctg 3628
Leu Thr Leu Gly Leu Met Leu Phe Tyr Arg Leu Val Phe Ile Leu
1105 1110 1115

cat gaa aca agc agg tcc act cag aaa gcc att gcc ttc tgt ctt 3673
His Glu Thr Ser Arg Ser Thr Gln Lys Ala Ile Ala Phe Cys Leu
1120 1125 1130

ggc tat ggc tgc cca ctt gcc atc tcg gtc atc acg ctg gga gcc 3718
Gly Tyr Gly Cys Pro Leu Ala Ile Ser Val Ile Thr Leu Gly Ala
1135 1140 1145

acc cag ccc cgg gaa gtc tat acg agg aag aat gtc tgt tgg ctc 3763
Thr Gln Pro Arg Glu Val Tyr Thr Arg Lys Asn Val Cys Trp Leu
1150 1155 1160

aac tgg gag gac acc aag gcc ctg ctg gct ttc gcc atc cca gca 3808
Asn Trp Glu Asp Thr Lys Ala Leu Leu Ala Phe Ala Ile Pro Ala
1165 1170 1175

ctg atc att gtg gtg gtg aac ata acc atc act att gtg gtc atc 3853
Leu Ile Ile Val Val Val Asn Ile Thr Ile Thr Ile Val Val Ile
1180 1185 1190

acc aag atc ctg agg cct tcc att gga gac aag cca tgc aag cag 3898
Thr Lys Ile Leu Arg Pro Ser Ile Gly Asp Lys Pro Cys Lys Gln
1195 1200 1205

gag aag agc agc ctg ttt cag atc agc aag agc att ggg gtc ctc 3943
Glu Lys Ser Ser Leu Phe Gln Ile Ser Lys Ser Ile Gly Val Leu
1210 1215 1220

aca cca ctc ttg ggc ctc act tgg ggt ttt ggt ctc acc act gtg 3988
Thr Pro Leu Leu Gly Leu Thr Trp Gly Phe Gly Leu Thr Thr Val
1225 1230 1235

ttc cca ggg acc aac ctt gtg ttc cat atc ata ttt gcc atc ctc 4033
Phe Pro Gly Thr Asn Leu Val Phe His Ile Ile Phe Ala Ile Leu
1240 1245 1250

aat gtc ttc cag gga tta ttc att tta ctc ttt gga tgc ctc tgg 4078
Asn Val Phe Gln Gly Leu Phe Ile Leu Leu Phe Gly Cys Leu Trp
1255 1260 1265

gat ctg aag gta cag gaa gct ttg ctg aat aag ttt tca ttg tcg 4123
Asp Leu Lys Val Gln Glu Ala Leu Leu Asn Lys Phe Ser Leu Ser
1270 1275 1280

aga tgg tct tca cag cac tca aag tca aca tcc ctg ggt tca tcc 4168
Arg Trp Ser Ser Gln His Ser Lys Ser Thr Ser Leu Gly Ser Ser
1285 1290 1295

aca cct gtg ttt tct atg agt tct cca ata tca agg aga ttt aac 4213
Thr Pro Val Phe Ser Met Ser Ser Pro Ile Ser Arg Arg Phe Asn
1300 1305 1310

aat ttg ttt ggt aaa aca gga acg tat aat gtt tcc acc cca gaa 4258
Asn Leu Phe Gly Lys Thr Gly Thr Tyr Asn Val Ser Thr Pro Glu
1315 1320 1325

gca acc agc tca tcc ctg gaa aac tca tcc agt gct tct tcg ttg 4303
Ala Thr Ser Ser Ser Leu Glu Asn Ser Ser Ser Ala Ser Ser Leu
1330 1335 1340

ctc aac taa gaacaggata atccaaccta cgtgacctcc cggggacagt 4352
Leu Asn
1345

ggctgtgctt ttaaaaagag atgcttgcaa agcaatgggg aacgtgttct cggggcaggt 4412

ttccgggagc agatgccaaa aagacttttt catagagaag aggctttctt ttgtaaagac 4472

agaataaaaa taattgttat gtttctgttt gttccctccc cctccccctt gtgtgatacc 4532

acatgtgtat agtatttaag tgaaactcaa gccctcaagg cccaacttct ctgtctatat 4592

tgtaatatag aatttcgaag agacattttc actttttaca cattgggcac aaagataagc 4652

tttgattaaa gtagtaagta aaaggctacc taggaaatac ttcagtgaat tctaagaagg 4712

aaggaaggaa gaaaggaagg aaagaaggga gggaaacagg gagaaaggga aaaagaagaa 4772

aaagagaaag atgaaaatag gaacaaataa agacaaacaa cattaagggc catattgtaa 4832

gatttccatg ttaatgatct aatataatca ctcagtgcaa cattgagaat tttttttaa 4892

tggctcaaaa atggaaactg aaagcaagtc atggggaatg aatactttgg gcagtatctt 4952

cctcatgtct tcttagctaa gaggaggaaa aaaaggctga aaaaataggg aggaaattcc 5012

ttcatcagaa cgacttcaag tggataacaa tatttataag aaatgaatgg aaggaaatat 5072

gatcctcctg agactaactt tgtatgttaa ggtttgaact aagtgaatgt atctgcagag 5132

gaagtattac aaagatatgt cattagatcc aagtgctgat taaattttta tagtttatca 5192

gaaaagcctt atattttagt ttgttccaca ttttgaaagc aaaaaatata tatttgatat 5252

acccttcaat tgccaaattt gatatgttgc actgaagaca gaccctgtca tatatttaat 5312

ggcttcaagc aggtacttct ctgtgcatta tagaatagat tttaataatc ttatagcatt 5372

gtatattatt attgctgttg tcactgttat tattattgtg gatactggcc cttggtgtgt 5432

tgcatagctc cctatgtatt ctctgtttcc atctttaagt tcccagacca atatacatta 5492

agagttttgc atggtctaaa ttgtgtttat tccaaccacg tggaaagctc ctggaaagaa 5552

attttacatt cggttgttct gtgctcctaa tgacacttga ccttgttgaa caaatggcag 5612

agcctttccc aaggatttga ttgtttgtga attatctgca tgtgtgcttt tttttggtgt 5672

gtatttcatt aaaaaatata aatatttatg aaaaaaaaaa aa 5714

<210> 61
<211> 1346
<212> PRT
<213> Homo sapiens

<400> 61

Met Lys Ser Pro Arg Arg Thr Thr Leu Cys Leu Met Phe Ile Val Ile
1 5 10 15

Tyr Ser Ser Lys Ala Ala Leu Asn Trp Asn Tyr Glu Ser Thr Ile His
20 25 30

Pro Leu Ser Leu His Glu His Glu Pro Ala Gly Glu Glu Ala Leu Arg
35 40 45

Gln Lys Arg Ala Val Ala Thr Lys Ser Pro Thr Ala Glu Glu Tyr Thr
50 55 60

Val Asn Ile Glu Ile Ser Phe Glu Asn Ala Ser Phe Leu Asp Pro Ile
65 70 75 80

Lys Ala Tyr Leu Asn Ser Leu Ser Phe Pro Ile His Gly Asn Asn Thr
85 90 95

Asp Gln Ile Thr Asp Ile Leu Ser Ile Asn Val Thr Thr Val Cys Arg
100 105 110

Pro Ala Gly Asn Glu Ile Trp Cys Ser Cys Glu Thr Gly Tyr Gly Trp
115 120 125

Pro Arg Glu Arg Cys Leu His Asn Leu Ile Cys Gln Glu Arg Asp Val
130 135 140

Phe Leu Pro Gly His His Cys Ser Cys Leu Lys Glu Leu Pro Pro Asn
145 150 155 160

Gly Pro Phe Cys Leu Leu Gln Glu Asp Val Thr Leu Asn Met Arg Val
165 170 175

Arg Leu Asn Val Gly Phe Gln Glu Asp Leu Met Asn Thr Ser Ser Ala
180 185 190

Leu Tyr Arg Ser Tyr Lys Thr Asp Leu Glu Thr Ala Phe Arg Lys Gly
195 200 205

Tyr Gly Ile Leu Pro Gly Phe Lys Gly Val Thr Val Thr Gly Phe Lys
210 215 220

Ser Gly Ser Val Val Val Thr Tyr Glu Val Lys Thr Thr Pro Pro Ser
225 230 235 240

Leu Glu Leu Ile His Lys Ala Asn Glu Gln Val Val Gln Ser Leu Asn
245 250 255

Gln Thr Tyr Lys Met Asp Tyr Asn Ser Phe Gln Ala Val Thr Ile Asn
260 265 270

Glu Ser Asn Phe Phe Val Thr Pro Glu Ile Ile Phe Glu Gly Asp Thr
275 280 285

Val Ser Leu Val Cys Glu Lys Glu Val Leu Ser Ser Asn Val Ser Trp
290 295 300

Arg Tyr Glu Glu Gln Gln Leu Glu Ile Gln Asn Ser Ser Arg Phe Ser
305 310 315 320

Ile Tyr Thr Ala Leu Phe Asn Asn Met Thr Ser Val Ser Lys Leu Thr
325 330 335

Ile His Asn Ile Thr Pro Gly Asp Ala Gly Glu Tyr Val Cys Lys Leu
340 345 350

Ile Leu Asp Ile Phe Glu Tyr Glu Cys Lys Lys Lys Ile Asp Val Met
355 360 365

Pro Ile Gln Ile Leu Ala Asn Glu Glu Met Lys Val Met Cys Asp Asn
370 375 380

Asn Pro Val Ser Leu Asn Cys Cys Ser Gln Gly Asn Val Asn Trp Ser
385 390 395 400

Lys Val Glu Trp Lys Gln Glu Gly Lys Ile Asn Ile Pro Gly Thr Pro
405 410 415

Glu Thr Asp Ile Asp Ser Ser Cys Ser Arg Tyr Thr Leu Lys Ala Asp
420 425 430

Gly Thr Gln Cys Pro Ser Gly Ser Ser Gly Thr Thr Val Ile Tyr Thr
435 440 445

Cys Glu Phe Ile Ser Ala Tyr Gly Ala Arg Gly Ser Ala Asn Ile Lys
450 455 460

Val Thr Phe Ile Ser Val Ala Asn Leu Thr Ile Thr Pro Asp Pro Ile
465 470 475 480

Ser Val Ser Glu Gly Gln Asn Phe Ser Ile Lys Cys Ile Ser Asp Val
485 490 495

Ser Asn Tyr Asp Glu Val Tyr Trp Asn Thr Ser Ala Gly Ile Lys Ile
500 505 510

Tyr Gln Arg Phe Tyr Thr Thr Arg Arg Tyr Leu Asp Gly Ala Glu Ser
515 520 525

Val Leu Thr Val Lys Thr Ser Thr Arg Glu Trp Asn Gly Thr Tyr His
530 535 540

Cys Ile Phe Arg Tyr Lys Asn Ser Tyr Ser Ile Ala Thr Lys Asp Val
545 550 555 560

Ile Val His Pro Leu Pro Leu Lys Leu Asn Ile Met Val Asp Pro Leu
565 570 575

Glu Ala Thr Val Ser Cys Ser Gly Ser His His Ile Lys Cys Cys Ile
580 585 590

Glu Glu Asp Gly Asp Tyr Lys Val Thr Phe His Met Gly Ser Ser Ser
595 600 605

Leu Pro Ala Ala Lys Glu Val Asn Lys Lys Gln Val Cys Tyr Lys His
610 615 620

Asn Phe Asn Ala Ser Ser Val Ser Trp Cys Ser Lys Thr Val Asp Val
625 630 635 640

Cys Cys His Phe Thr Asn Ala Ala Asn Asn Ser Val Trp Ser Pro Ser
645 650 655

Met Lys Leu Asn Leu Val Pro Gly Glu Asn Ile Thr Cys Gln Asp Pro
660 665 670

Val Ile Gly Val Gly Glu Pro Gly Lys Val Ile Gln Lys Leu Cys Arg
675 680 685

Phe Ser Asn Val Pro Ser Ser Pro Glu Ser Pro Ile Gly Gly Thr Ile
690 695 700

Thr Tyr Lys Cys Val Gly Ser Gln Trp Glu Glu Lys Arg Asn Asp Cys
705 710 715 720

Ile Ser Ala Pro Ile Asn Ser Leu Leu Gln Met Ala Lys Ala Leu Ile
725 730 735

Lys Ser Pro Ser Gln Asp Glu Met Leu Pro Thr Tyr Leu Lys Asp Leu
740 745 750

Ser Ile Ser Ile Asp Lys Ala Glu His Glu Ile Ser Ser Ser Pro Gly
755 760 765

Ser Leu Gly Ala Ile Ile Asn Ile Leu Asp Leu Leu Ser Thr Val Pro
770 775 780

Thr Gln Val Asn Ser Glu Met Met Thr His Val Leu Ser Thr Val Asn
785 790 795 800

Val Ile Leu Gly Lys Pro Val Leu Asn Thr Trp Lys Val Leu Gln Gln
805 810 815

Gln Trp Thr Asn Gln Ser Ser Gln Leu Leu His Ser Val Glu Arg Phe
820 825 830

Ser Gln Ala Leu Gln Ser Gly Asp Ser Pro Pro Leu Ser Phe Ser Gln

835 840 845

Thr Asn Val Gln Met Ser Ser Thr Val Ile Lys Ser Ser His Pro Glu
850 855 860

Thr Tyr Gln Gln Arg Phe Val Phe Pro Tyr Phe Asp Leu Trp Gly Asn
865 870 875 880

Val Val Ile Asp Lys Ser Tyr Leu Glu Asn Leu Gln Ser Asp Ser Ser
885 890 895

Ile Val Thr Met Ala Phe Pro Thr Leu Gln Ala Ile Leu Ala Gln Asp
900 905 910

Ile Gln Glu Asn Asn Phe Ala Glu Ser Leu Val Met Thr Thr Thr Val
915 920 925

Ser His Asn Thr Thr Met Pro Phe Arg Ile Ser Met Thr Phe Lys Asn
930 935 940

Asn Ser Pro Ser Gly Gly Glu Thr Lys Cys Val Phe Trp Asn Phe Arg
945 950 955 960

Leu Ala Asn Asn Thr Gly Gly Trp Asp Ser Ser Gly Cys Tyr Val Glu
965 970 975

Glu Gly Asp Gly Asp Asn Val Thr Cys Ile Cys Asp His Leu Thr Ser
980 985 990

Phe Ser Ile Leu Met Ser Pro Asp Ser Pro Asp Pro Ser Ser Leu Leu
995 1000 1005

Gly Ile Leu Leu Asp Ile Ile Ser Tyr Val Gly Val Gly Phe Ser
1010 1015 1020

Ile Leu Ser Leu Ala Ala Cys Leu Val Val Glu Ala Val Val Trp
1025 1030 1035

Lys Ser Val Thr Lys Asn Arg Thr Ser Tyr Met Arg His Thr Cys
1040 1045 1050

Ile Val Asn Ile Ala Ala Ser Leu Leu Val Ala Asn Thr Trp Phe
1055 1060 1065

Ile Val Val Ala Ala Ile Gln Asp Asn Arg Tyr Ile Leu Cys Lys
1070 1075 1080

Thr Ala Cys Val Ala Ala Thr Phe Phe Ile His Phe Phe Tyr Leu
1085 1090 1095

Ser Val Phe Phe Trp Met Leu Thr Leu Gly Leu Met Leu Phe Tyr
1100 1105 1110

Arg Leu Val Phe Ile Leu His Glu Thr Ser Arg Ser Thr Gln Lys
1115 1120 1125

Ala Ile Ala Phe Cys Leu Gly Tyr Gly Cys Pro Leu Ala Ile Ser
1130 1135 1140

Val Ile Thr Leu Gly Ala Thr Gln Pro Arg Glu Val Tyr Thr Arg
1145 1150 1155

Lys Asn Val Cys Trp Leu Asn Trp Glu Asp Thr Lys Ala Leu Leu
1160 1165 1170

Ala Phe Ala Ile Pro Ala Leu Ile Ile Val Val Val Asn Ile Thr
1175 1180 1185

Ile Thr Ile Val Val Ile Thr Lys Ile Leu Arg Pro Ser Ile Gly
1190 1195 1200

Asp Lys Pro Cys Lys Gln Glu Lys Ser Ser Leu Phe Gln Ile Ser
1205 1210 1215

Lys Ser Ile Gly Val Leu Thr Pro Leu Leu Gly Leu Thr Trp Gly
1220 1225 1230

Phe Gly Leu Thr Thr Val Phe Pro Gly Thr Asn Leu Val Phe His
1235 1240 1245

Ile Ile Phe Ala Ile Leu Asn Val Phe Gln Gly Leu Phe Ile Leu
1250 1255 1260

Leu Phe Gly Cys Leu Trp Asp Leu Lys Val Gln Glu Ala Leu Leu
1265 1270 1275

Asn Lys Phe Ser Leu Ser Arg Trp Ser Ser Gln His Ser Lys Ser
1280 1285 1290

Thr Ser Leu Gly Ser Ser Thr Pro Val Phe Ser Met Ser Ser Pro
1295 1300 1305

Ile Ser Arg Arg Phe Asn Asn Leu Phe Gly Lys Thr Gly Thr Tyr
1310 1315 1320

Asn Val Ser Thr Pro Glu Ala Thr Ser Ser Ser Leu Glu Asn Ser
1325 1330 1335

Ser Ser Ala Ser Ser Leu Leu Asn
1340 1345

<210> 62
<211> 2282
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1257)..(2189)

<400> 62
tttacctaaa tgttcattta tttaaacttg aaaatgtgac tgcaaaagga gtcaagaagt 60

ttagtactca acacatttct tctttgtaga aataactgct caagaagata aattgttacc 120

taacagttga ttgatgaata tgaaatcaga gcaagagggg ctaattagag attttactag 180

ctagggctat ttacccaaac catctatcca gactgtagac atagaatcac cagatggcta 240

ggatcctggt gcagctggtt ccccttttc cttaccctga atgtcattaa ggatgcattg 300

ccaaatgctg cccctctggc ctgatgacta cactccatat tggtcacctg ccttcttttt 360

cctacagtct tcctccagac aggcacgcca tacaaccgac tacacttggg ctcactgaat 420

gaatcacatt cttctgctgt gcctcccaga gatttcatca aagcacccgc agtggccttc 480

tggaggctcc tcaaactctc cactcaatgt ttcctgaaag tgtcgtgttt ctcacctcta 540

gattgttatt ctcatcagtt acatgtgggt ttcacaaatt tatttctcag aatgcaagtc 600

tgtctcttat atcctcggga aacacacctt tatatcccag ttagtactga caaaaattaa 660

actagggact ggccaaaaac agtgcctttc ctcactttaa tctcactaaa gtagataaga 720

ctcaagttat tttgttcttg caatggcatt gacaaatgtt tgcaccaaaa accatgttga 780

agttcattaa ggaaactgtg atccaagatc caaggtcaaa aaaacaaatt catcaattca 840

gcacaccacc aactcacagg ctaagcatct tactgctaat tcattgatgc tgccatttgt 900

caagtgccaa attgaattat tgatttgtca ataatttcct tctgttggtt acttatatag 960

tatattgcaa ttcttgttgc tgaagtcagc tacacttttt ctatttgaaa aacaatttct 1020

tgcatttggg atttcaggta tagtgattgt tacaaatatg aaggacttga attaacagca 1080

agttttcaag taaaacttta cttatgtata actgaatgag ttcttaaaga catttactaa 1140

caattttcca caaactaaaa atttataaaa caataaataa aatagacttt aaaaaaaagc 1200

gtgtcacaca gctgcttgtt ttttgtttgt ttctttgttt gtttttttagt agtgaa atg 1259
Met
1

```
gtg aaa aat cag aca atg gtc aca gag ttc ctc cta ctg gga ttt ctc     1307
Val Lys Asn Gln Thr Met Val Thr Glu Phe Leu Leu Leu Gly Phe Leu
            5                   10                  15

ctg ggc cca agg att cag atg ctc ctc ttt ggg ctc ttc tcc ctg ttc     1355
Leu Gly Pro Arg Ile Gln Met Leu Leu Phe Gly Leu Phe Ser Leu Phe
        20                  25                  30

tat gtc ttc acc ctg ctg ggg aat ggg acc atc ctg ggg ctc atc tca     1403
Tyr Val Phe Thr Leu Leu Gly Asn Gly Thr Ile Leu Gly Leu Ile Ser
    35                  40                  45

ctg gac tcc aga ctc cac acc ccc atg tac ttc ttc ctc tca cac ctg     1451
Leu Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ser His Leu
50                  55                  60                  65

gcc gtc gtc aac atc gcc tat gcc tgc aac aca gtg ccc cag atg ctg     1499
Ala Val Val Asn Ile Ala Tyr Ala Cys Asn Thr Val Pro Gln Met Leu
                70                  75                  80

gtg aac ctc ctg cat cca gcc aag ccc atc tcc ttt gct ggc tgc atg     1547
Val Asn Leu Leu His Pro Ala Lys Pro Ile Ser Phe Ala Gly Cys Met
            85                  90                  95

aca cag acc ttt ctc ttt ttg agt ttt gca cat act gaa tgc ctc ctg     1595
Thr Gln Thr Phe Leu Phe Leu Ser Phe Ala His Thr Glu Cys Leu Leu
        100                 105                 110

ttg gtg ctg atg tcc tac gat cgg tac gtg gcc atc tgc cac cct ctc     1643
Leu Val Leu Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Leu
    115                 120                 125

cga tat ttc atc atc atg acc tgg aaa gtc tgc atc act ctg gcc atc     1691
Arg Tyr Phe Ile Ile Met Thr Trp Lys Val Cys Ile Thr Leu Ala Ile
130                 135                 140                 145

act tcc tgg aca tgt ggc tcc ctc ctg gct atg gtc cat gtg agc ctc     1739
Thr Ser Trp Thr Cys Gly Ser Leu Leu Ala Met Val His Val Ser Leu
                150                 155                 160

atc cta aga ctg ccc ttt tgt ggg cct cgt gaa atc aac cac ttc ttc     1787
Ile Leu Arg Leu Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe Phe
            165                 170                 175

tgt gaa atc ctg tct gtc ctc agg ctg gcc tgt gct gat acc tgg ctc     1835
Cys Glu Ile Leu Ser Val Leu Arg Leu Ala Cys Ala Asp Thr Trp Leu
        180                 185                 190

aac cag gtg gtc atc ttt gca gcc tgc atg ttc atc ctg gtg gga cca     1883
Asn Gln Val Val Ile Phe Ala Ala Cys Met Phe Ile Leu Val Gly Pro
    195                 200                 205

ctc tgc ctg gtg ctg gtc tcc tac tca cac atc ctg gcg gcc atc ctg     1931
Leu Cys Leu Val Leu Val Ser Tyr Ser His Ile Leu Ala Ala Ile Leu
210                 215                 220                 225

agg atc cag tct ggg gag ggc cgc aga aag gcc ttc tcc acc tgc tcc     1979
Arg Ile Gln Ser Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys Ser
                230                 235                 240

tcc cac ctc tgc gta gtg gga ctc ttc ttt ggc agc gcc atc gtc atg     2027
Ser His Leu Cys Val Val Gly Leu Phe Phe Gly Ser Ala Ile Val Met
            245                 250                 255
```

tac atg gcc cct aag tcc cgc cat cct gag gag cag cag aag gtc ctt 2075
Tyr Met Ala Pro Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val Leu
260 265 270

ttt cta ttt tac agt tct ttc aac ccg atg cta aac ccc ctg att tac 2123
Phe Leu Phe Tyr Ser Ser Phe Asn Pro Met Leu Asn Pro Leu Ile Tyr
275 280 285

aac ctg agg aat gta gag gtc aag ggt gcc ctg agg aga gca ctg tgc 2171
Asn Leu Arg Asn Val Glu Val Lys Gly Ala Leu Arg Arg Ala Leu Cys
290 295 300 305

aag gaa agt cat tcc taa gaggtgtgac atttgaactg ccagcctcag 2219
Lys Glu Ser His Ser
310

ttgtcacgtg gactctttga tgcccaatta tttgcctcaa tccagaaaag tttacttctt 2279

ctt 2282

<210> 63
<211> 310
<212> PRT
<213> Homo sapiens

<400> 63

Met Val Lys Asn Gln Thr Met Val Thr Glu Phe Leu Leu Leu Gly Phe
1 5 10 15

Leu Leu Gly Pro Arg Ile Gln Met Leu Leu Phe Gly Leu Phe Ser Leu
20 25 30

Phe Tyr Val Phe Thr Leu Leu Gly Asn Gly Thr Ile Leu Gly Leu Ile
35 40 45

Ser Leu Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ser His
50 55 60

Leu Ala Val Val Asn Ile Ala Tyr Ala Cys Asn Thr Val Pro Gln Met
65 70 75 80

Leu Val Asn Leu Leu His Pro Ala Lys Pro Ile Ser Phe Ala Gly Cys
85 90 95

Met Thr Gln Thr Phe Leu Phe Leu Ser Phe Ala His Thr Glu Cys Leu
100 105 110

Leu Leu Val Leu Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro
115 120 125

Leu Arg Tyr Phe Ile Ile Met Thr Trp Lys Val Cys Ile Thr Leu Ala
130 135 140

Ile Thr Ser Trp Thr Cys Gly Ser Leu Leu Ala Met Val His Val Ser
145 150 155 160

Leu Ile Leu Arg Leu Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe
165 170 175

Phe Cys Glu Ile Leu Ser Val Leu Arg Leu Ala Cys Ala Asp Thr Trp
180 185 190

Leu Asn Gln Val Val Ile Phe Ala Ala Cys Met Phe Ile Leu Val Gly
195 200 205

Pro Leu Cys Leu Val Leu Val Ser Tyr Ser His Ile Leu Ala Ala Ile
210 215 220

Leu Arg Ile Gln Ser Gly Glu Gly Arg Arg Lys Ala Phe Ser Thr Cys
225 230 235 240

Ser Ser His Leu Cys Val Val Gly Leu Phe Phe Gly Ser Ala Ile Val
245 250 255

Met Tyr Met Ala Pro Lys Ser Arg His Pro Glu Glu Gln Gln Lys Val
260 265 270

Leu Phe Leu Phe Tyr Ser Ser Phe Asn Pro Met Leu Asn Pro Leu Ile
275 280 285

Tyr Asn Leu Arg Asn Val Glu Val Lys Gly Ala Leu Arg Arg Ala Leu
290 295 300

Cys Lys Glu Ser His Ser
305 310